



TREASURE DATA

# 글로벌 사례로 보는 데이터로 돈 버는 법



고영혁

트레저데이터 데이터사이언티스트/아키텍트  
dylan@treasure-data.com







# TREASURE DATA 소개



# 트레저데이터 회사 개요

Treasure Data 는 서비스 형태로 데이터 분석 인프라를 제공하는 세계 선도 기업으로서, 데이터 파이프라인과 데이터 플랫폼 관리의 각종 복잡한 이슈를 해결하여, 조직이 데이터로부터 가치를 추출하는 데에만 온전히 집중할 수 있도록 합니다.

- 설립연도: 2011년 12월
- 위치: Mountain View, CA (HQ), Tokyo, Seoul
- 투자현황: \$47M raised from top Silicon Valley VCs (Now Series C)
- 개발하여 후원하는 글로벌 유명 오픈소스 프로젝트 목록

	fluentd	• Very popular data collector
	MessagePack	• Data serializer
	embulk	• Enterprise data bulk loader
	Fluent Bit	• Data collector from IoT/Embedded devices

## Treasure Data Investors



**Jerry Yang**  
Yahoo! Founder



**Bill Tai**  
ex-CRV, leading angel investor



**James Lindenbaum**  
Heroku Founder

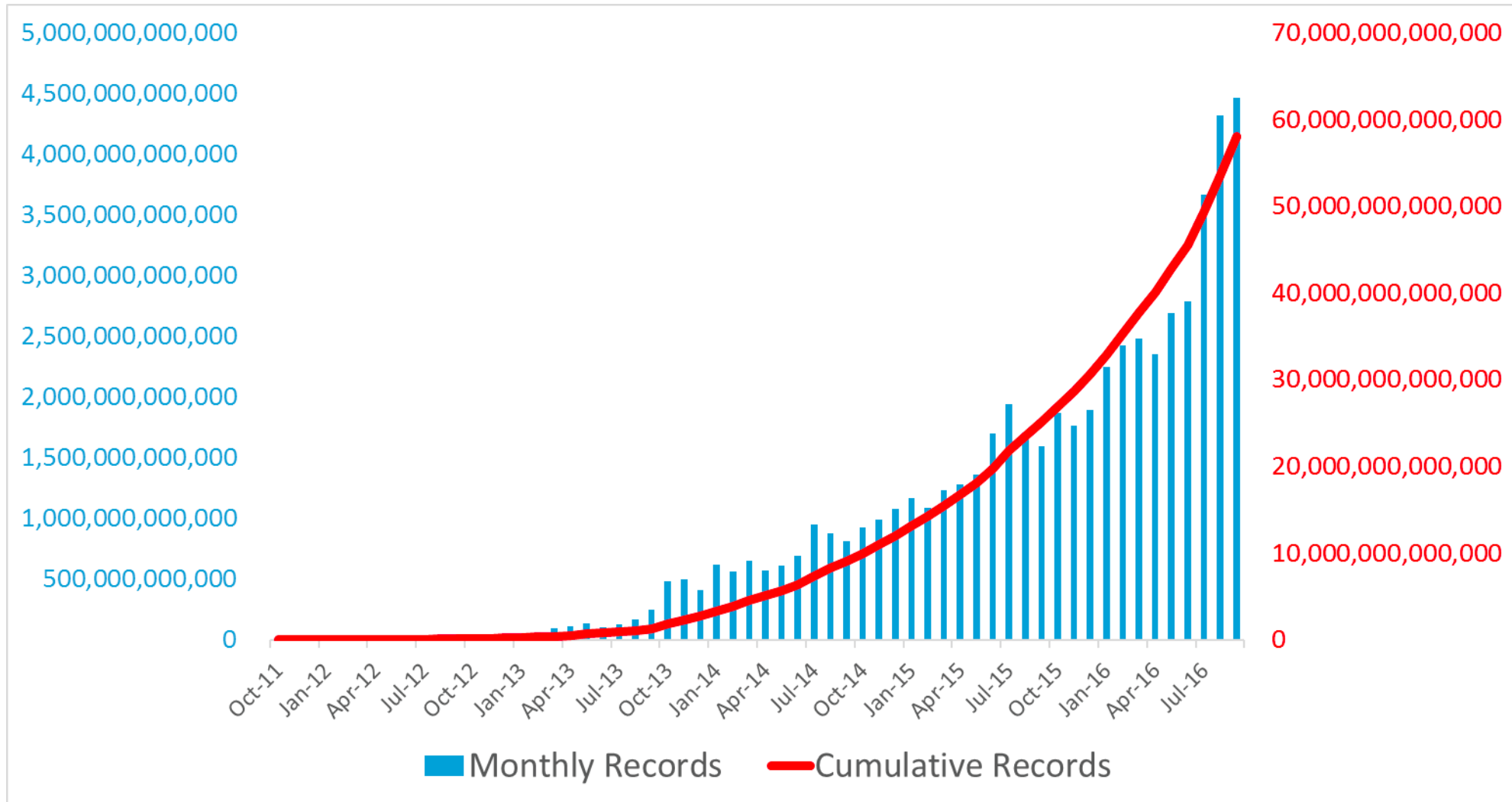
## Treasure Data's Data Collection Technology Users





# The Growth of Treasure Data

**60,000,000,000,000+** Rows (October 2016)





# 트레저데이터 창업자 소개

Hiro Yoshikawa

CEO



- 와세다 대학 재학 중에 Red Hat, Inc.에서 근무. 엔지니어로 경력을 시작하고 교육사업, 제품 마케팅, 사업개발을 담당. 2009년에 미쓰이물산의 벤처 캐피탈 사업인 "미쓰이 벤처스"의 멤버로 미국으로 감.
- 오픈소스 소프트웨어 기업에 대한 투자와 경영에 종사한 경험을 살려 2011년에 미국 트레저데이터사를 창업하여 최고 경영자(CEO)로 취임, 현재에 이르고 있다.
- Red Hat에서의 오픈소스 소프트웨어 사업 경험과 미국에서의 사업 경험을 살려, 트레저데이터의 경영 전반 책임자로 활약중.

Kazuki Ota

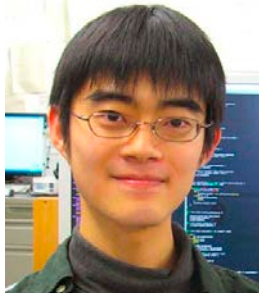
CTO



- 고등학교 시절부터 Linux의 데스크탑 환경 KDE 커미터로 OSS 커뮤니티에 공헌함.
- 2006년, 도쿄대학 이학부 정보과학과 재학중에 주식회사 Preferred Infrastructure의 최고기술책임자(CTO)로 취임. 여기에서 Hadoop을 만나 2009년에 "Hadoop 사용자 모임 재팬"을 설립. 현재 등록자수는 1,700명을 넘어 세계 최대규모의 Hadoop 커뮤니티가 되어있다.
- 2011년에 퇴직 후, 미국에서 트레저데이터사를 창업하고 최고기술책임자(CTO)로 취임, 현재에 이르고 있다.

Sada Furuhashi

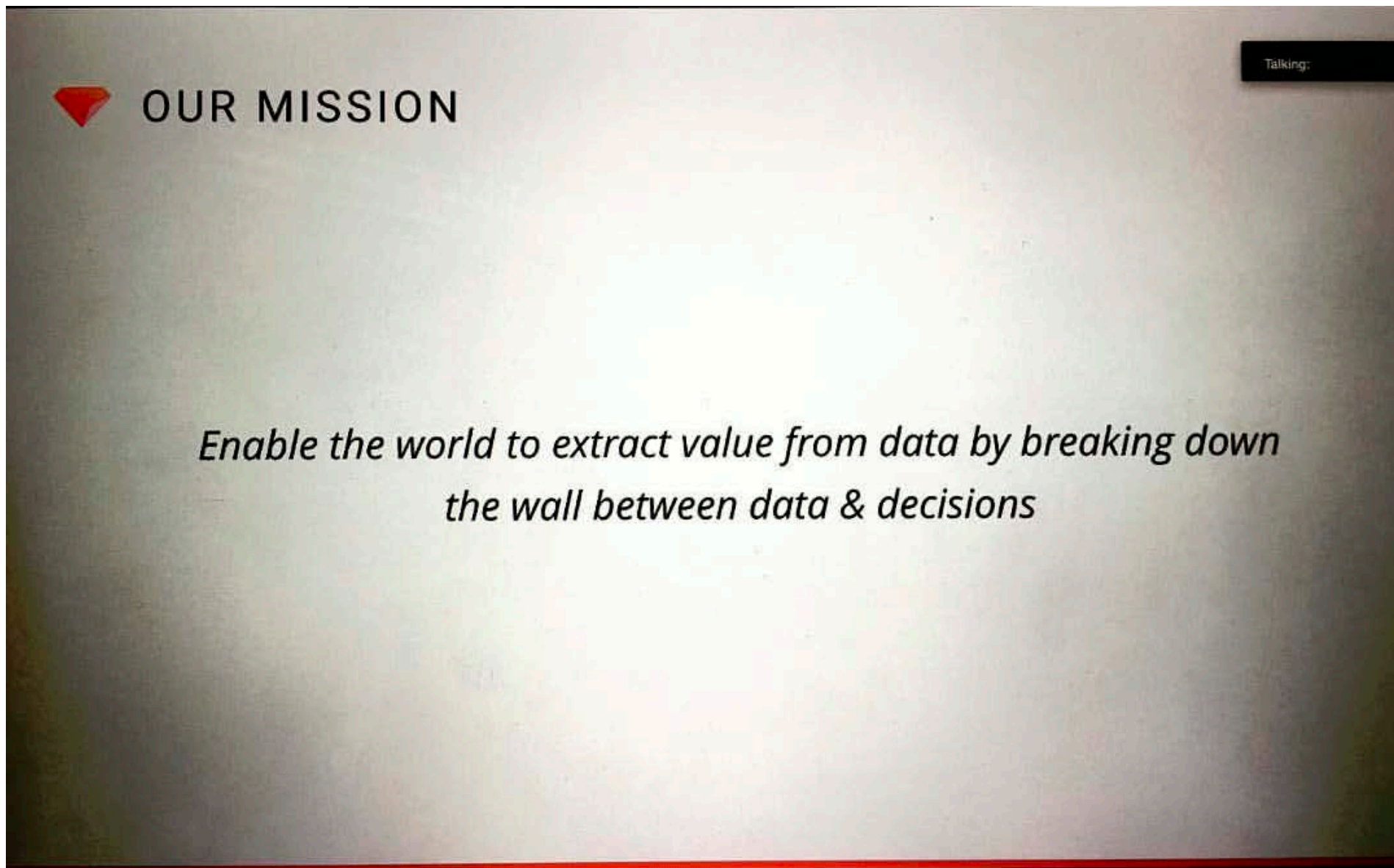
Chief Software Architect



- 2012년 쓰쿠바대학 대학원을 졸업. 학창시절인 2010년에 일본OSS추진포럼에서 "일본OSS장려상"을 수상.
- 또, 다국어 통신이 가능한 분산 인프라기술 "MessagePack(메시지 팩)"을 개발. 2008년에 오픈소스 소프트웨어로 공개하여 현재는 전 세계적으로 이용되고 있다.
- 2011년에는 미국 트레저데이터사를 요시카와, 오타와 함께 창업. Hadoop기반의 클라우드 데이터 분석 플랫폼의 구축 및 구현을 담당하고있다. 2011년부터 로그 컬렉터 "Fluentd"의 개발을 시작.
- 현재는 Web서비스나 광고, CDN등을 비롯한 수많은 산업 분야에서 전 세계적으로 도입되어 2014년에는 Google Cloud Platform의 로그 컬렉터로 표준적으로 채용됨. 2015년 "Embulk"의 개발을 시작.



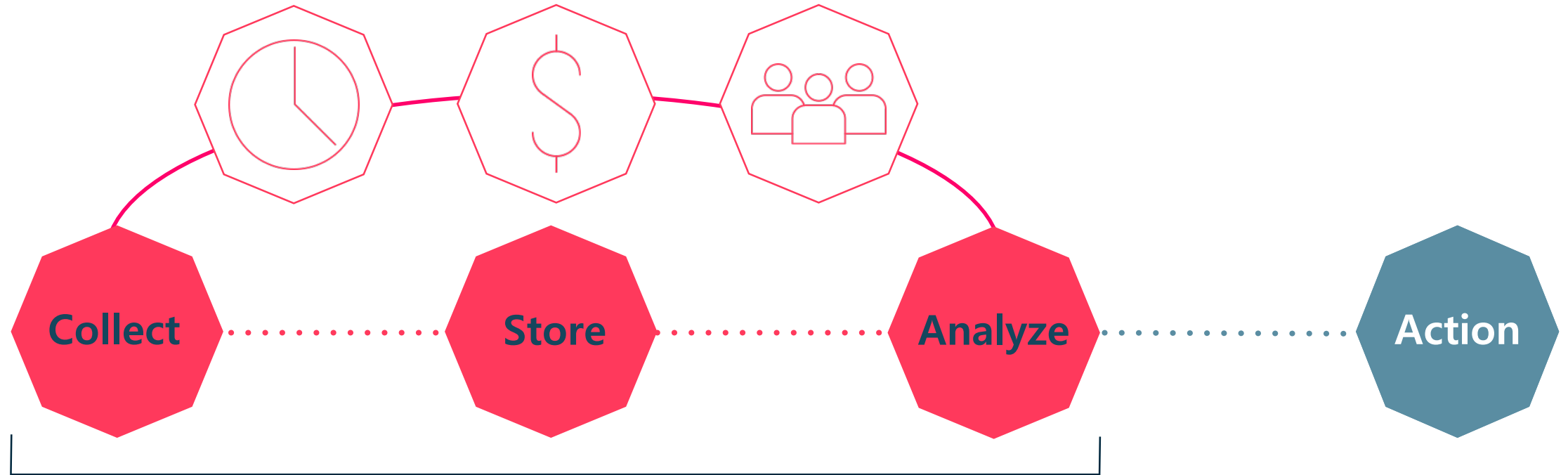
# 트레저데이터의 미션



## OUR MISSION

*Enable the world to extract value from data by breaking down  
the wall between data & decisions*

# ◆ 데이터가 돈이 되기 위한 기본문제와 트레저데이터의 해결 접근법

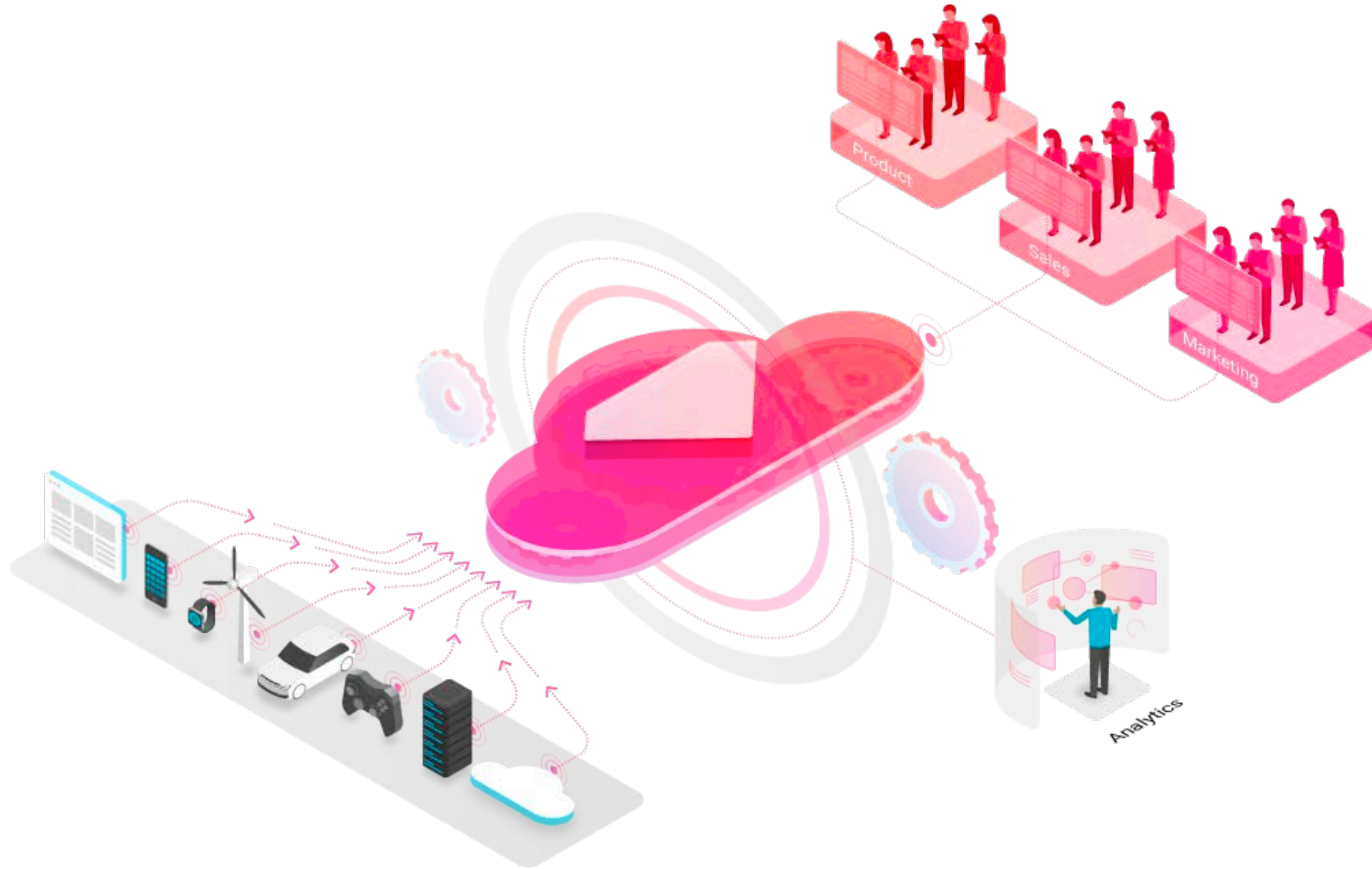


Essential Infra for "Data Driven Action"

## Treasure Data

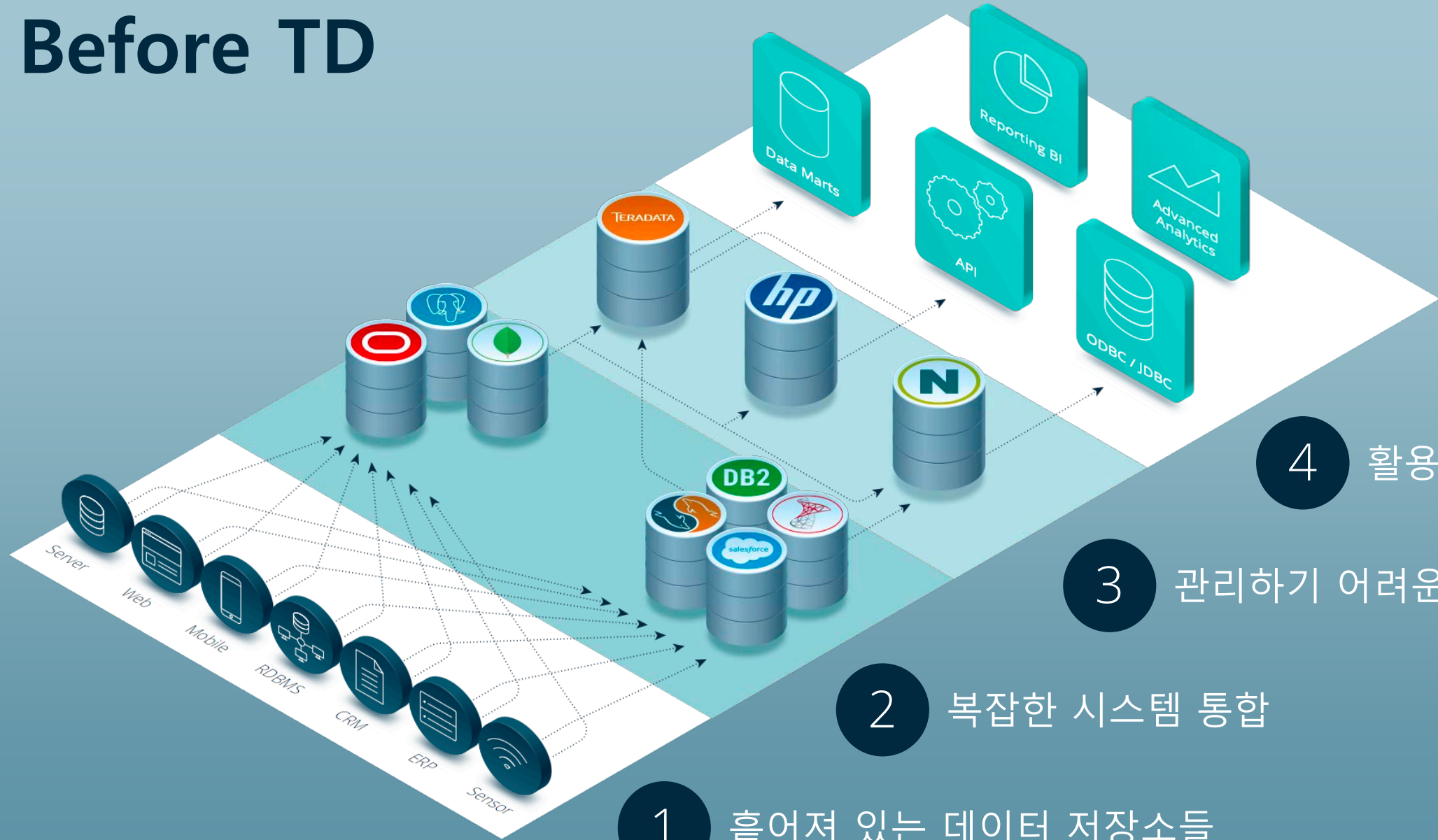
- Cloud Data Storage and Interface
- Fully Managed SaaS
- Holistic Live Data Management Platform
- Enable to Make Data Automation (Data Product) Easily and cost efficiently

# ◆ 데이터가 돈이 되기 위한 기본문제와 트레저데이터의 해결 접근법





# Before TD



1 흩어져 있는 데이터 저장소들

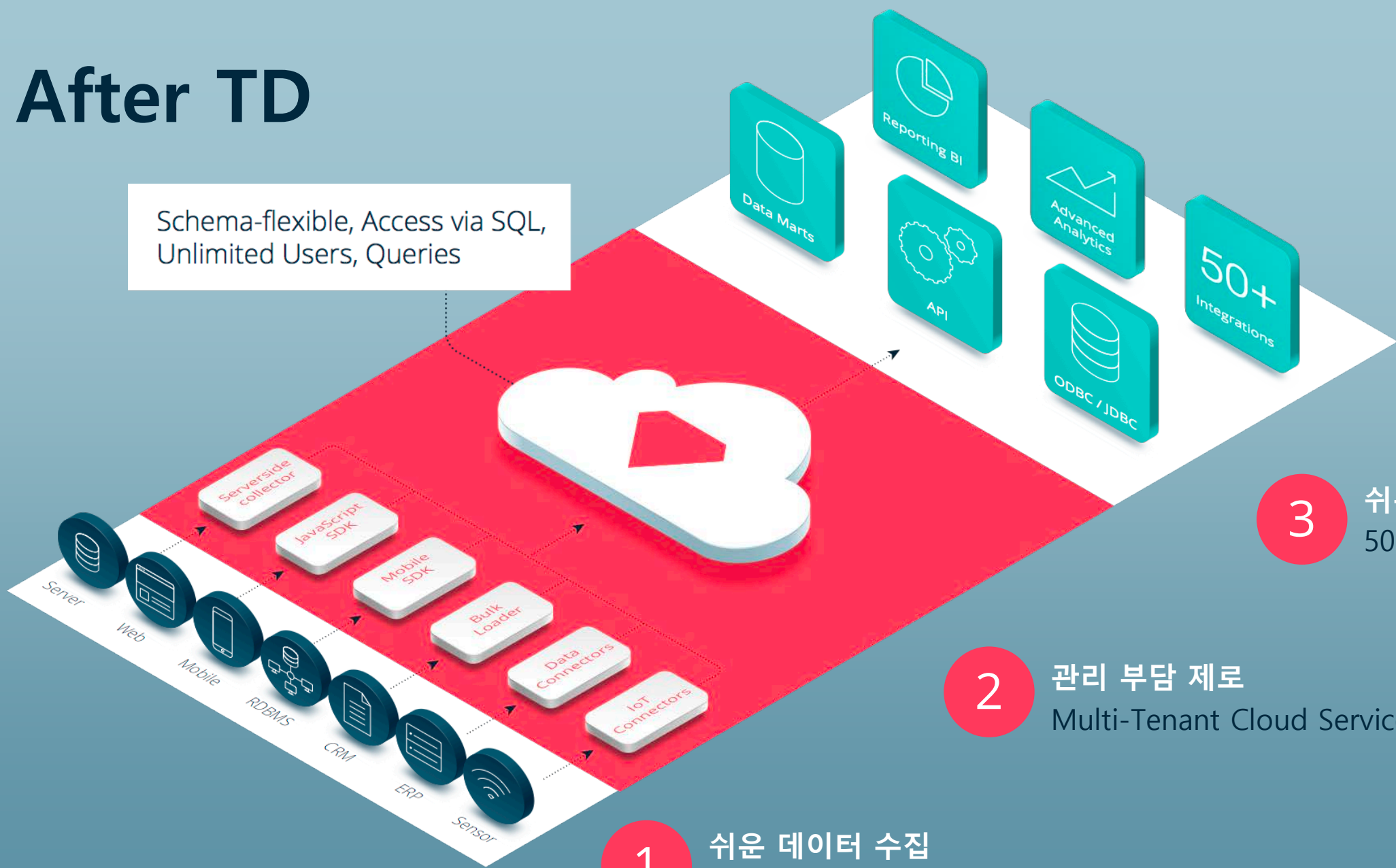
2 복잡한 시스템 통합

3 관리하기 어려운 분석계

4 활용 안되는 KPIs

# After TD

Schema-flexible, Access via SQL,  
Unlimited Users, Queries

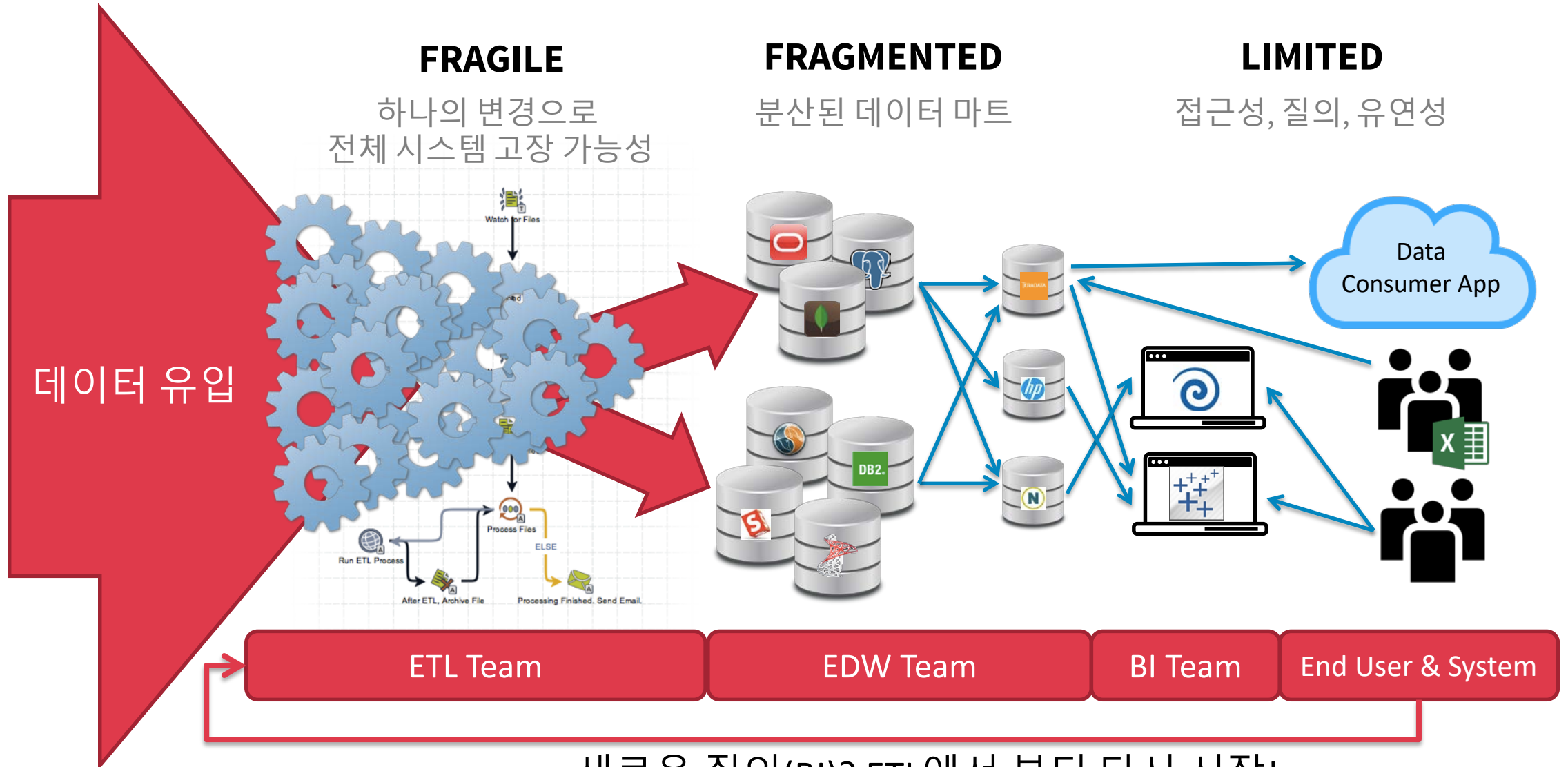


**1** 쉬운 데이터 수집  
300+ Data Sources

**2** 관리 부담 제로  
Multi-Tenant Cloud Service

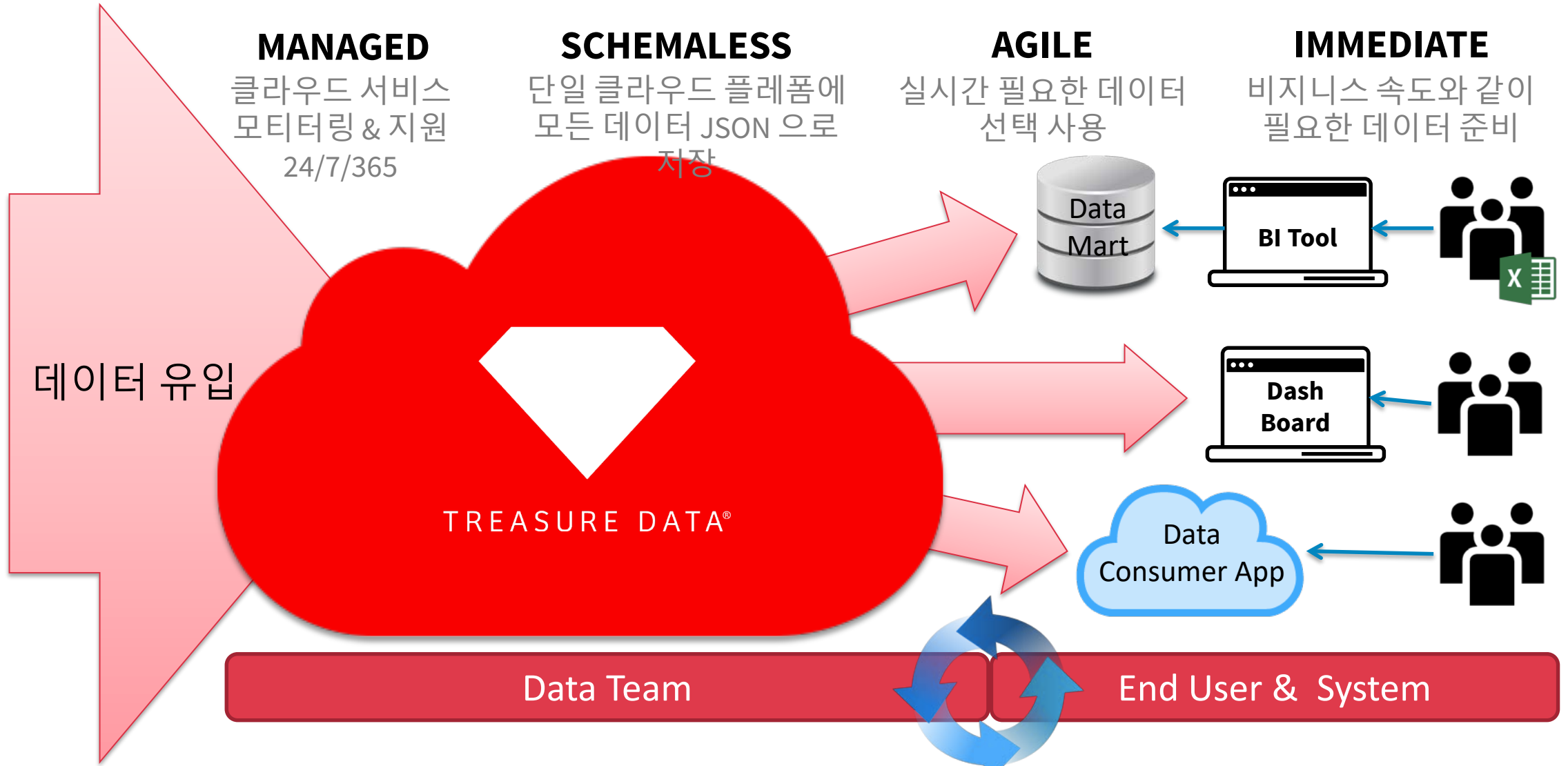
**3** 쉬운 솔루션 연계  
50+ Data Outputs

# Traditional ETL/BI Process



새로운 질의(BI)? ETL에서 부터 다시 시작!

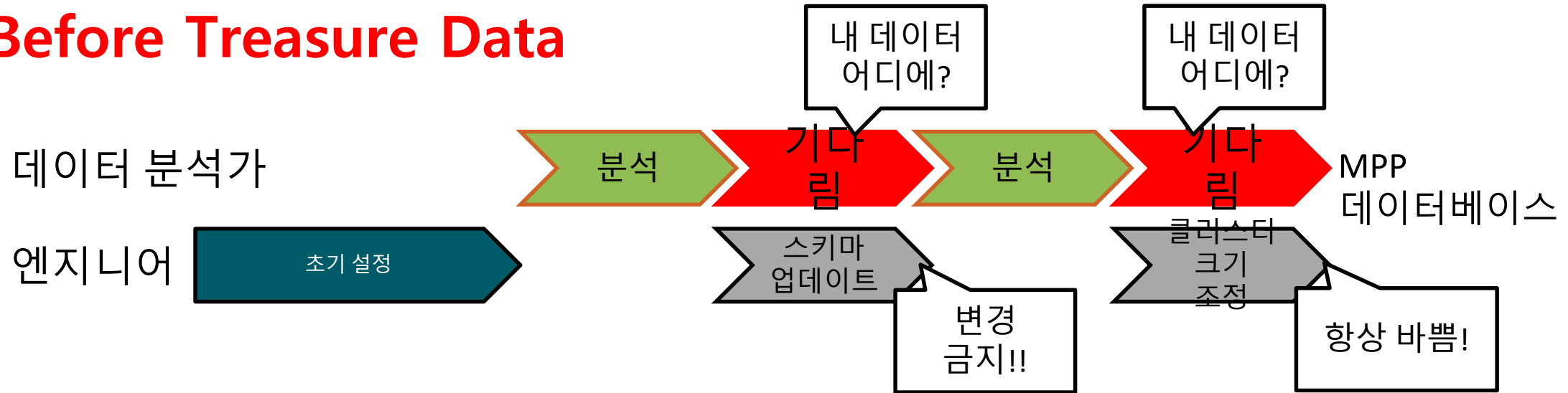
# ◆ A Completely Different Data Approach



모든데이터 저장, 필요한 데이터 선택 사용.

# ◆ A Completely Different Data Approach

## Before Treasure Data



데이터 분석가

계획, 분석 그리고 행위!

트레저 데이터는  
스키마 적용이  
자유로운(Schema  
Less) 클라우드  
서비스

엔지니어

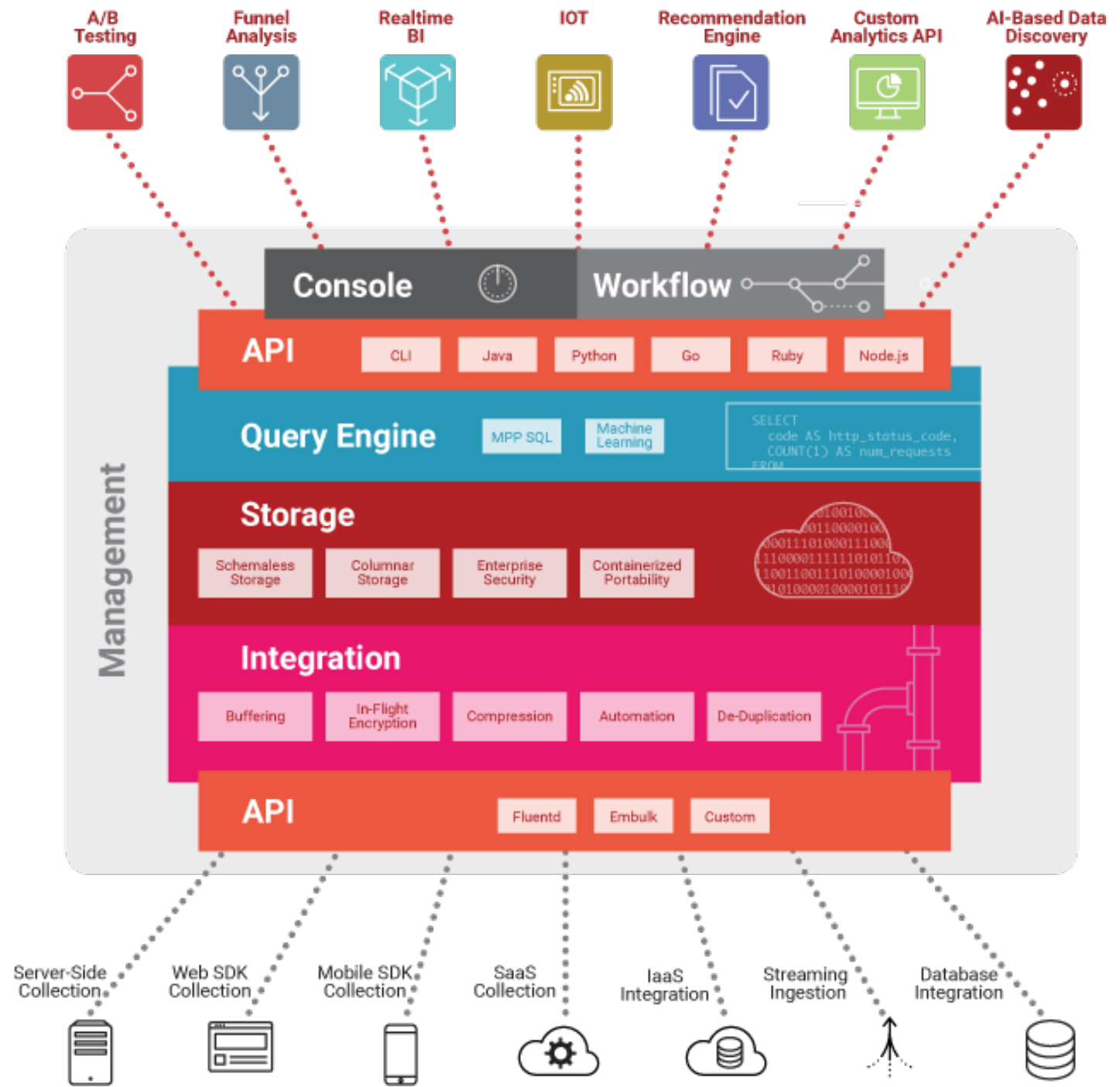
초기 설정

개발에 집중!

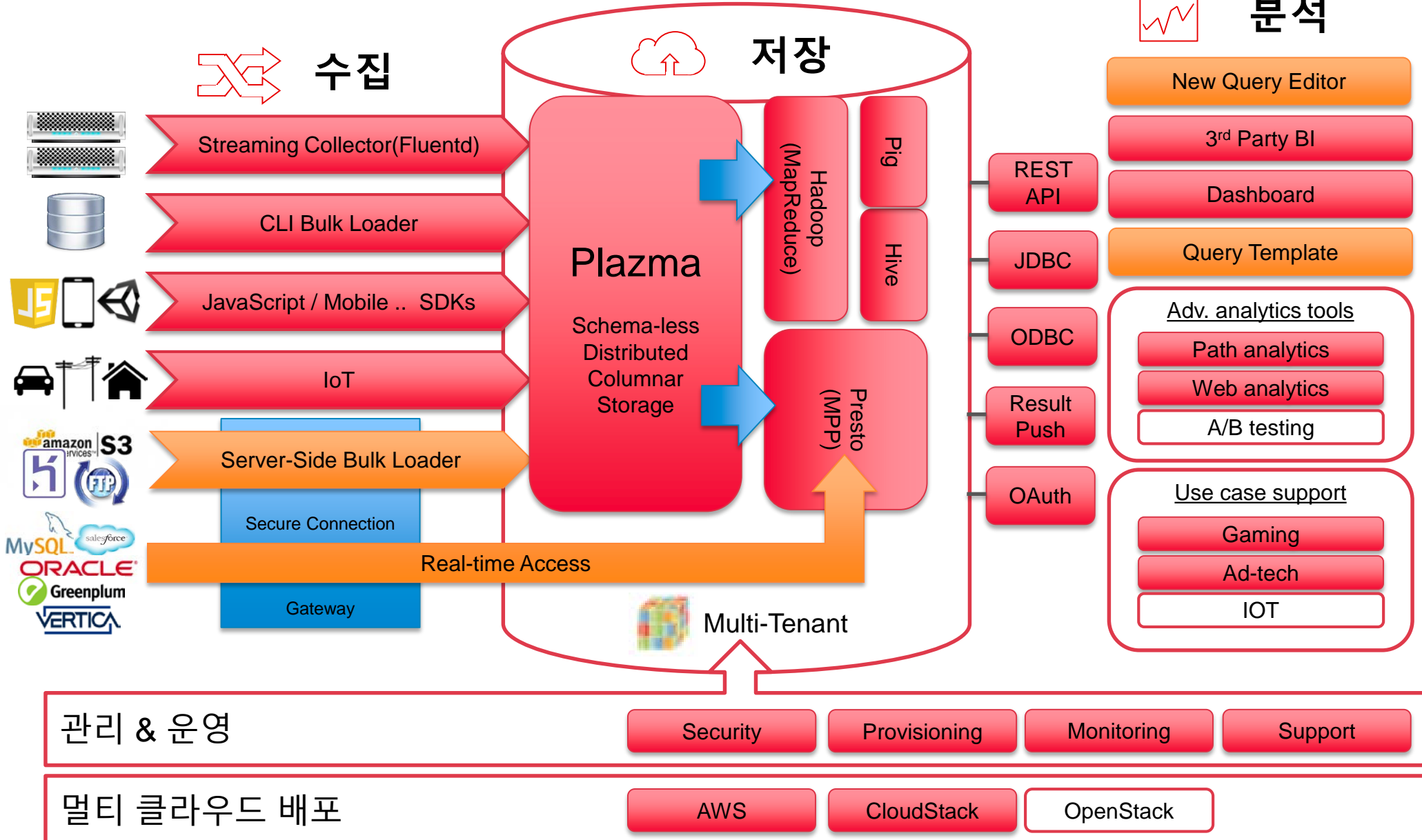
## After Treasure Data



# SERVICE ARCHITECTURE



# Technical Architecture



# ◆ 데이터 생태계의 혁신을 만들어낸 트레저데이터의 핵심 기술 역량

## Data Import



공동창업자 후루하시가 만든 스트리밍 형식의 로그 수집 도구  
Google, Microsoft, AWS 와 같은 글로벌 기업들에서 로그 수집 도구로 사용되고 있음



공동창업자 후루하시가 만든 벌크 형식의 로그 수집 도구  
오픈소스 소프트웨어로 공개한 지 1년 반만에 100개 이상의 플러그인이 개발되었으며, WEB 서비스 기업을 중심으로 사용되고 있음



트레저데이터가 개발한 IoT · 임베디드 기반을 위한 스트리밍 형식의 로그 수집 도구  
C언어로 개발되고 있어서 가볍고, 의존성에 구애 받지 않고 이용 가능

## Data Processing



트레저데이터가 개발한 프로그래밍 필요 없이 머신러닝을 실행할 수 있는 SQL 쿼리 형태의 라이브러리. 일본 주요 애드테크 기업들에서 CTR 예측 등에 활용



대규모 데이터 병렬 분산처리 기반  
오타 CTO 가 세계 최대규모의 Hadoop 일본 유저 그룹을 설립



Facebook 이 개발한 병렬 분산처리 기반 오픈소스 소프트웨어 프로젝트로서, Facebook, Teradata 와 함께 트레저데이터도 개발에 참여

## Storage



공동창업자 후루하시가 개발한 바이너리 기반의 고속 변환이 가능한 데이터 압축 포맷  
Pinterest, Uber 등에서도 데이터 저장용 포맷으로 사용하고 있음



트레저데이터가 독자 개발한 타임스탬프를 인덱스로 하는 스키마리스 컬럼기반 데이터베이스. 시계열 로그 빅데이터 처리에 특히 강함





# 솔루션 비교

## On-Premise



## In-Cloud



## Managed Service



	On-Premise	In-Cloud	Managed Service
운영해야 할 인프라	하드웨어, 소프트웨어	시간당 인스턴스	없음
투입할 직원 리소스	 구축 / 유지	 설정 / 모니터링	 최초 설정 한 번
확장성 & 민첩성	제한됨 / 비쌈	제한됨 / 비쌈	최저 비용 무한 자유도 (플렉서블 스키마 기반)
지원	자체 팀 필요	별도 비용 지불	기본 포함
활용까지 소요 기간	10-12 개월	4-6 개월	2-4 주
총비용	\$\$\$\$\$	\$\$\$	\$



# 다양한 분야의 글로벌 초우량 고객사

## Digital Marketing

*EC*

*Media*

*Game/SNS*

*Retail*

*Finance*

*Maker*

*Telecommunication*

*Technology*

## AD Tech

*Agency / Trading Desk*

























*DMP / DSP*

*Ad-Network*

## IoT



# 트레저데이터의 기술 (Fluentd, MessagePack, Embulk)을 사용하는 대표 기업



# 데이터로 돈 버는 글로벌 사례 #1

MUJI : 전통적 리테일에서 데이터 기반 O2O

## ◆ MUJI (무인양품) 기본 소개

- 1980년 회사 설립
- 연매출 22억 USD
- 연수익 2억 USD
- 매장 현황 - 일본 401개 / 해외 301개
- MUJI = NO brand name or designer name
- **트레저데이터 활용 - 2013년부터**

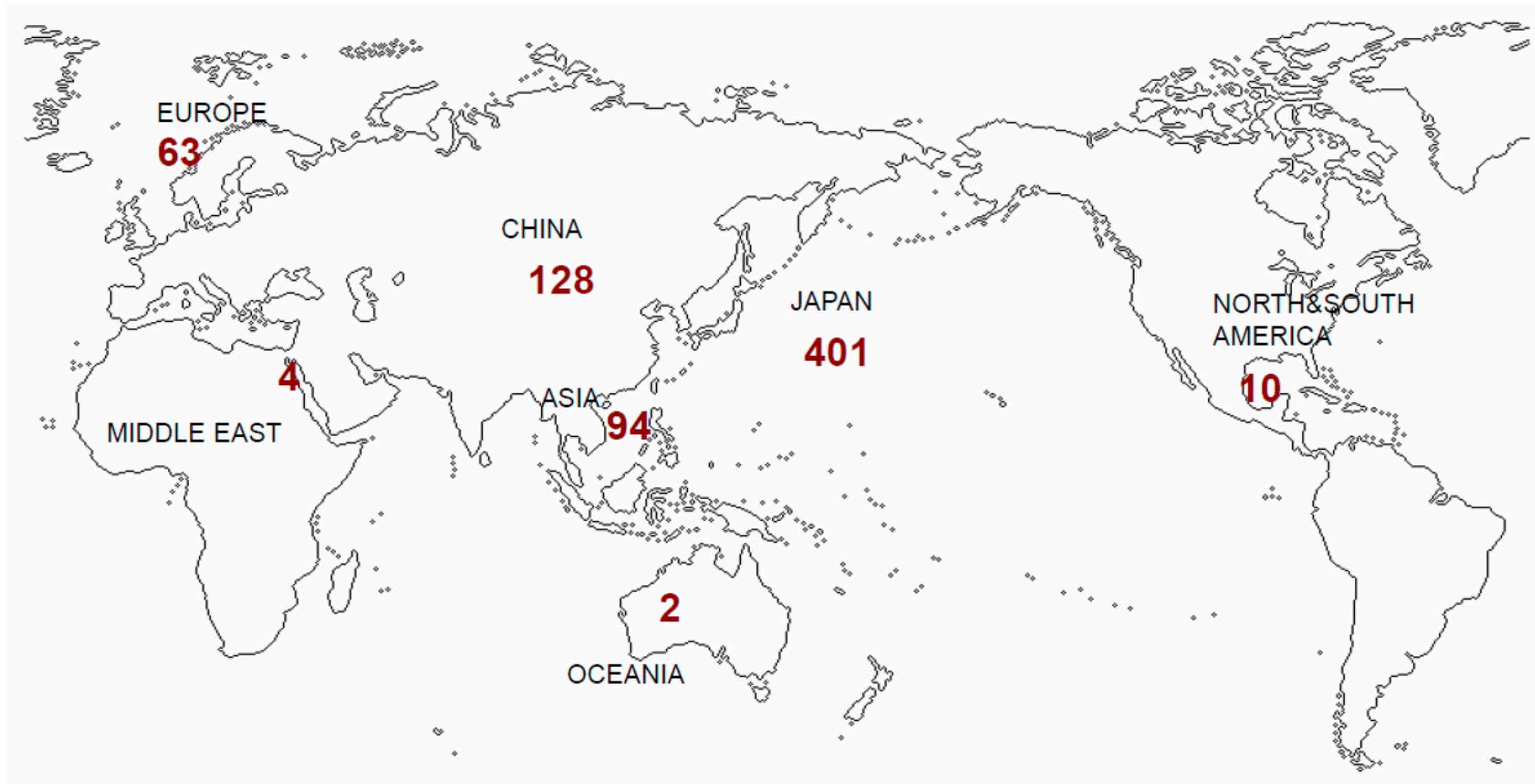
※ 2015년 2월 기준 자료

# MUJI

## 無印良品



# ◆ MUJI (무인양품) 기본 소개





# MUJI (무인양품) 기본 소개





## MUJI – Needs & Challenge (Pain Points)

### ❖ NEEDS

- 매출의 성장
- 전세계에 걸쳐 있는 다수 매장의 관리
- 고객 로열티의 관리
- 데이터를 활용하는 것이 근본적인 답이라는 것은 알고 있음

### ❖ CHALLENGE

- 방대한 웹/앱/매장 데이터의 통합 문제
- 기존 시스템을 최대한 건드리지 않고 패러다임 시프트
- 시간. 시간. 시간
- Scalability

Treasure Data  
로 해결





# 무인양품의 O2O 기반 리테일 혁신의 HUB – MUJI Passport



앱 첫화면



ID



매장 체크인



관련 뉴스



제품 검색

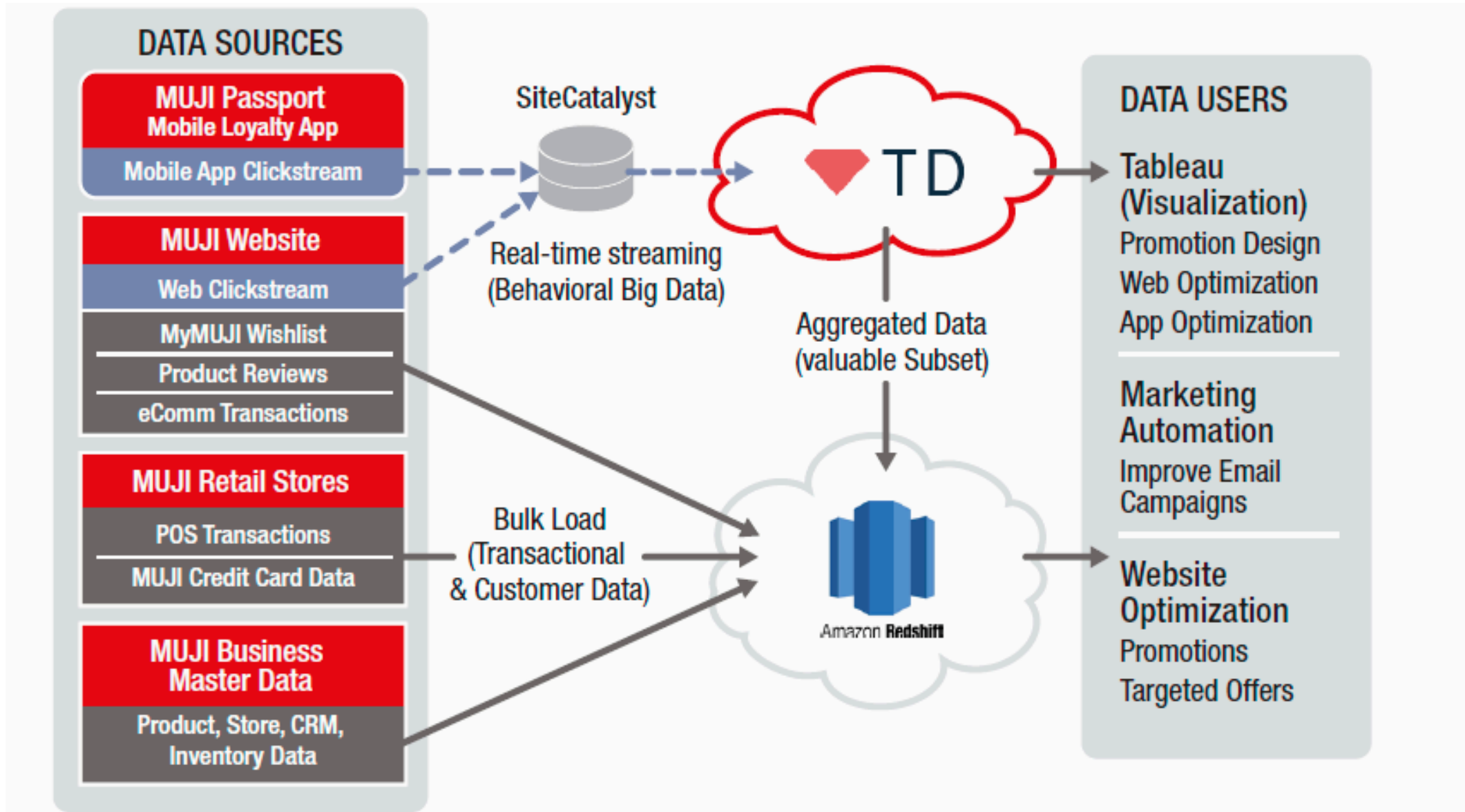


매장 안내

- 바코드 기반의 고객 ID 기능
- MUJI 쇼핑 포인트(마일리지) 기능
- 혜택 쿠폰 발급
- 매장 체크인 기능
- 관련 뉴스
- 쇼핑 가이드 – 제품 검색, 매장 안내
- 선호도/관여도 정보 수집 (Like, Have)
- MUJI 마일리지의 통합 및 보너스  
마일리지를 활용한 ID/정보 통합 유도 –  
소셜 ID, 온라인 사이트 ID, 카드 ID

**Digital ID + Owned Media + Store & Product Search**

# ◆ MUJI – Treasure data 활용 O2O Analytics Platform



## MUJI – Treasure data 활용 핵심 및 결과

### ❖ 활용 핵심

- 고객의 온라인 행동과 오프라인 행동 패턴을 데이터를 통해 모두 파악
- 온라인에서의 행동과 매장의 재고 및 매출 현황을 토대로, 고객의 매장 방문 시점에서 해당 고객 및 매출에 최적화된 스펙의 쿠폰을 실시간으로 생성하여 사용 유도

### ❖ 활용 결과

- 매출 **46%** 증가
- 전체 매장 평균 쿠폰 활용률 **100%** 증가
- 오프라인 매장 내 체류시간 및 동선량 증가



## MUJI – Next Step

- 현재 TD를 통해 연 30억 트랜잭션 이상의 데이터를 저비용으로 기존 시스템 건드리지 않고 효율적으로 처리
- 이렇게 쌓인 데이터를 TD의 Machine-Learning-As-a-Service 인 HiveMall 을 이용해서 매장 내 위치 기반 실시간 프로모션 엔진 준비
- TD 인터페이스를 통한 다양한 연계로 프로모션/마케팅 자동화를 효율적 구축



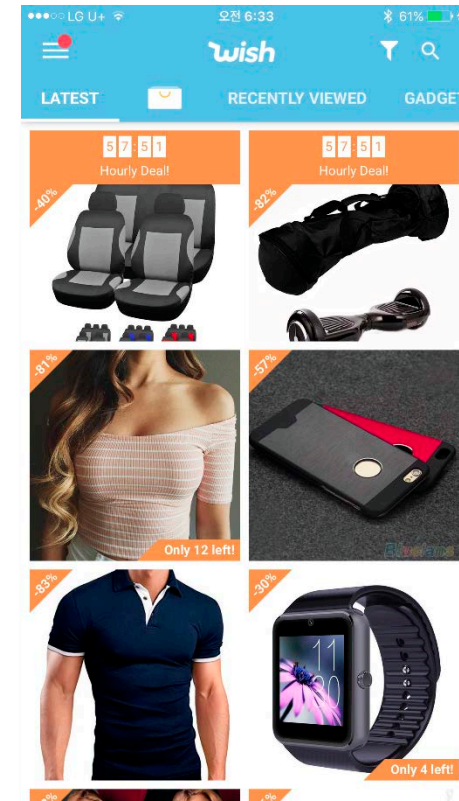
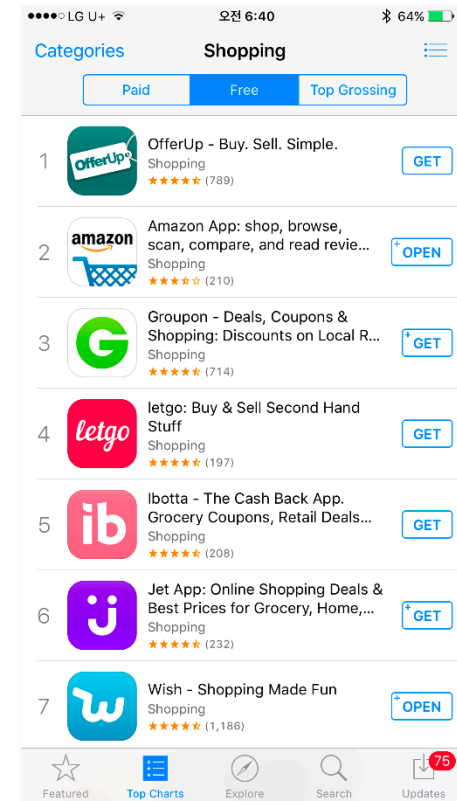
## 데이터로 돈 버는 글로벌 사례 #2

**WISH : 개인화&자동화를 통한 쇼핑 최적화**

## ◆ WISH – 기본 소개

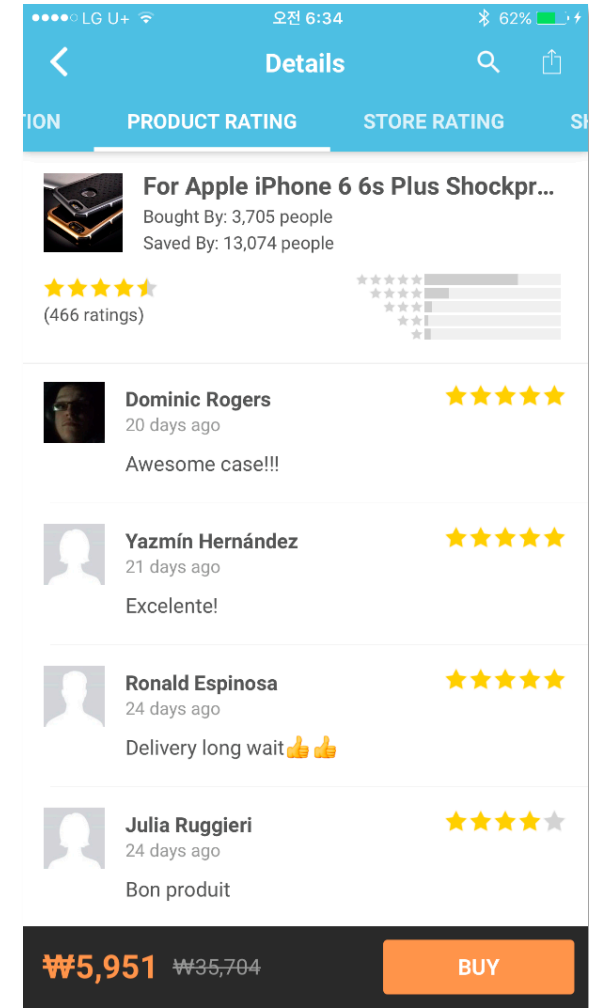
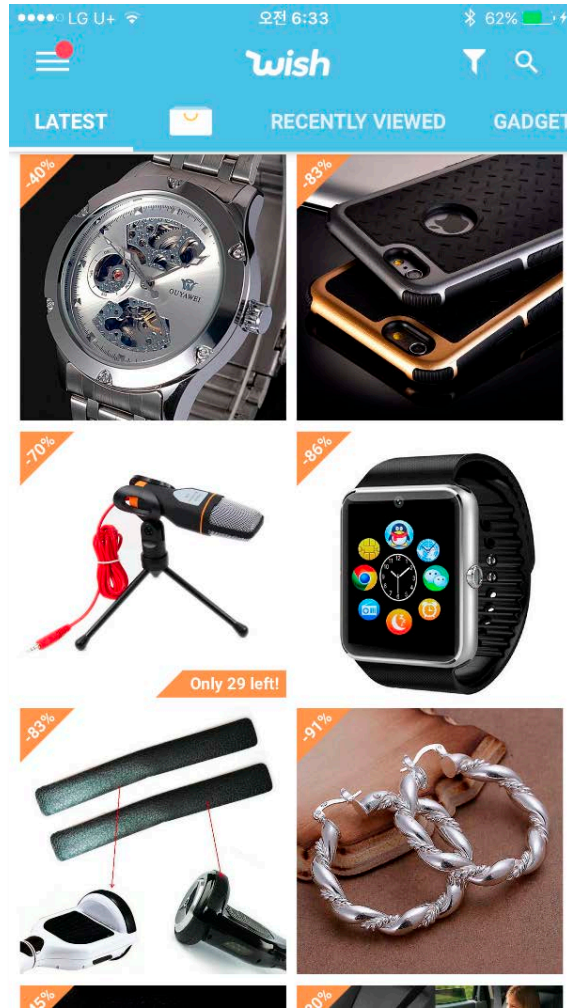
- 개인화 추천 기반 모바일 커머스
- 야후, 구글, 페이스북 엔지니어가 2011년 창업
- 창업 후 18개월만에 1500만 일활성화 이용자
- 전세계에서 두번째로 많이 다운로드된 쇼핑앱
- 현재 전세계 2억 일활성화 이용자
- 2015년 말 기업가치 \$30억 (유니콘~데카콘)
- 2012년부터 TD 고객. 추후 TD에 투자까지

# wish



# ◆ WISH – 기본 소개

- 개인화 추천 기반 모바일 커머스. 80~90% 까지 할인 상품 다수





## WISH – Needs & Challenge (Pain Points)

### ❖ NEEDS

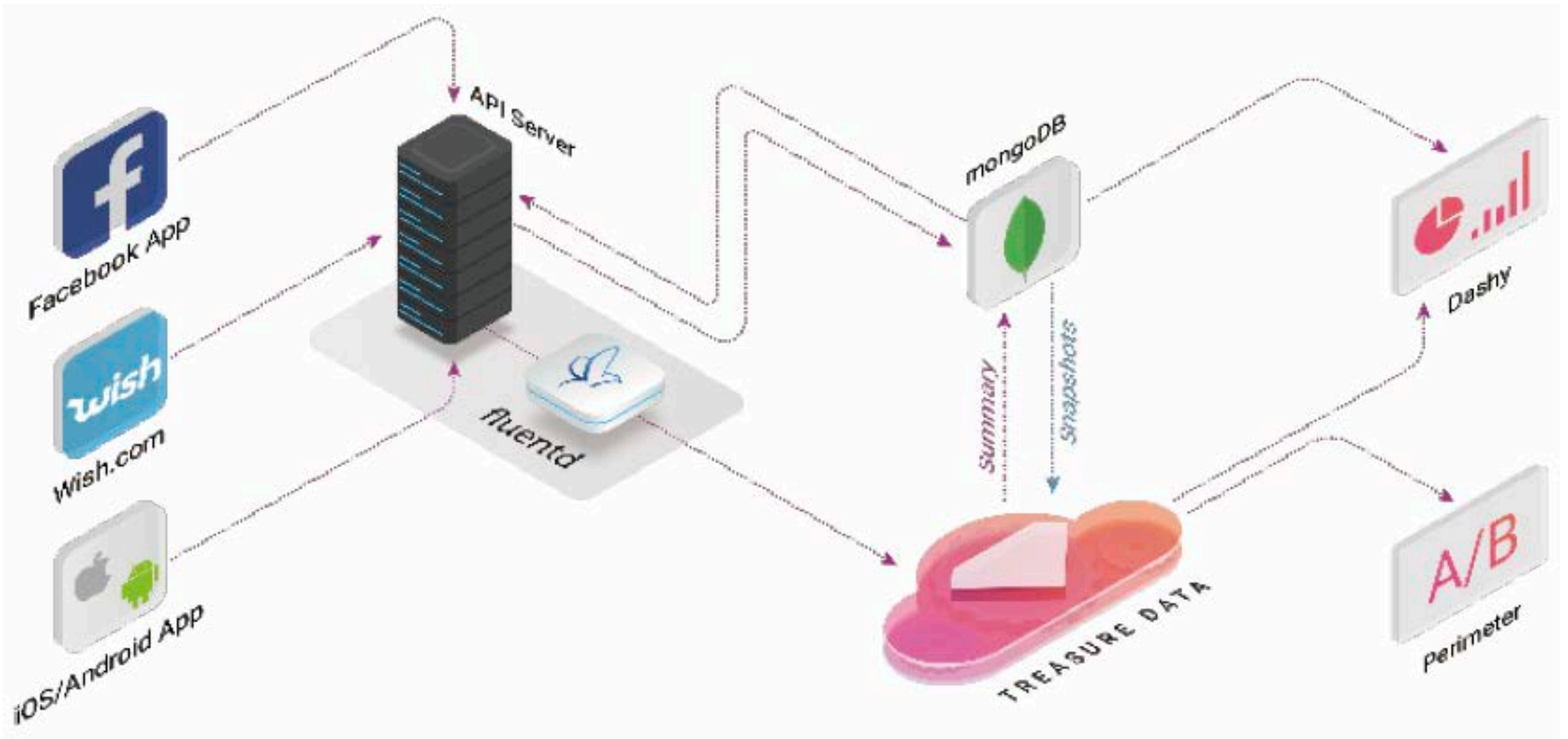
- 핵심 고객 세그먼트인 개인화된 경험 원하는 여성에 대한 정확한 이해
- 비즈니스에 핵심이 되는 액션을 빨리 수행
- 전 직원간의 유기적인 데이터 기반 커뮤니케이션과 액션

### ❖ CHALLENGE

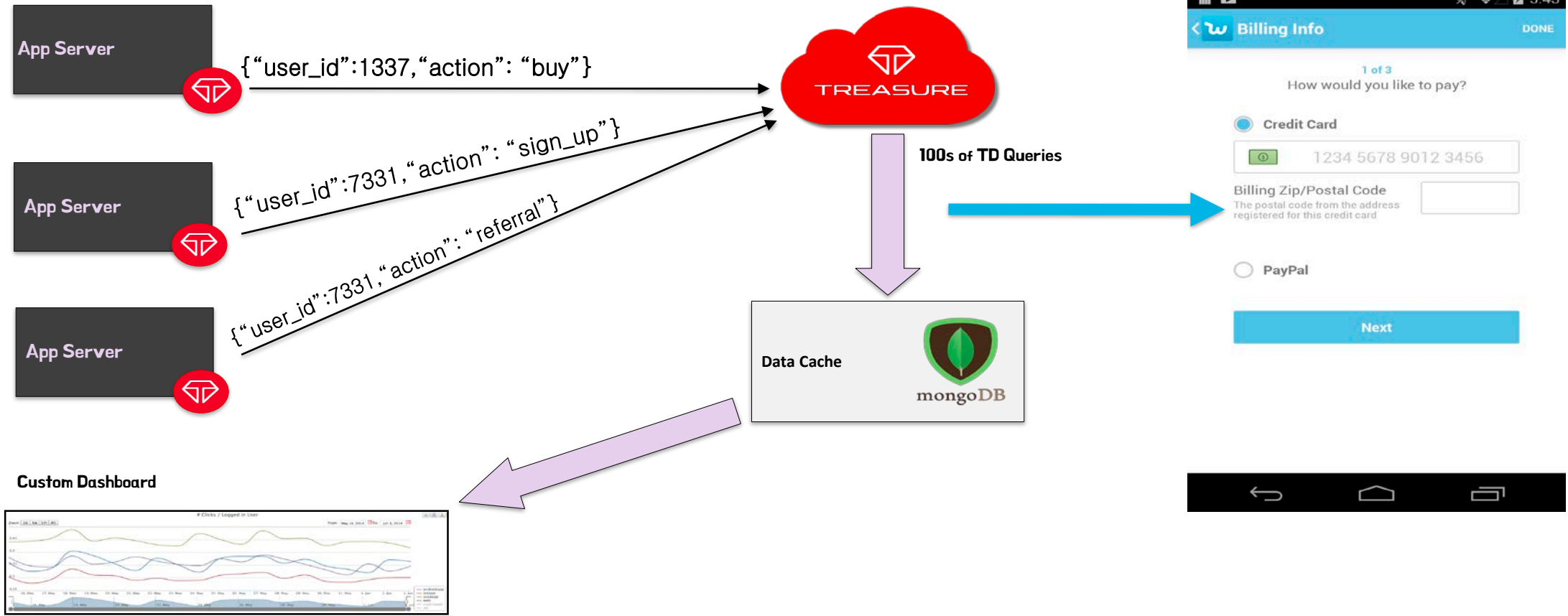
- 위시의 대부분은 여성을 잘 모르는 남성 엔지니어 (남탕)
- 구글, 야후, 페이스북 등에서 온 데이터, 머신러닝 전문가들이지만 이커머스에 필요한 방대한 것들을 다 알지도 못하고, 절대적인 사람 숫자도 부족



# ❖ WISH – Treasure Data 활용 아키텍처



# WISH – Treasure Data 활용 A/B 테스트 시스템



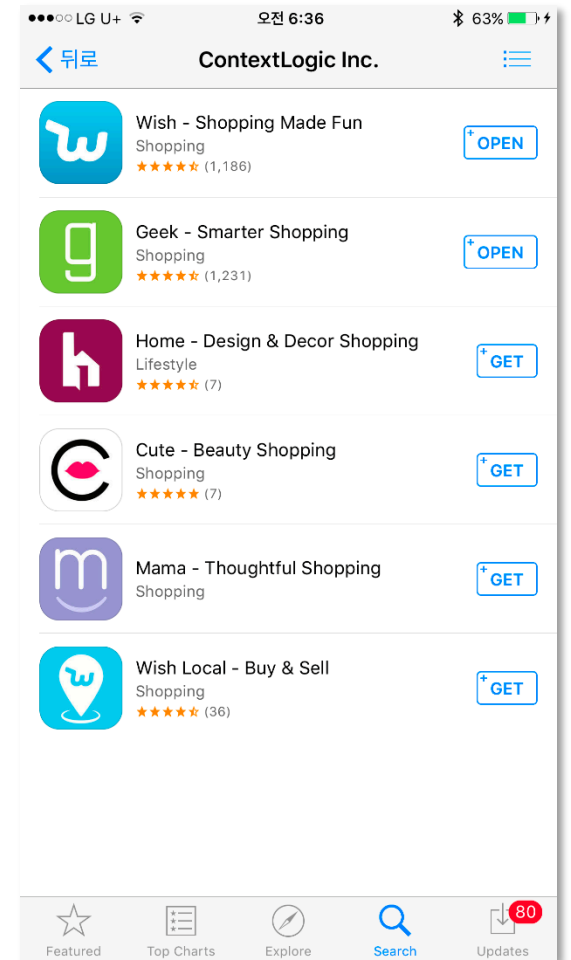
# ❖ WISH – Treasure Data 활용 핵심 및 결과

## ❖ 활용 핵심

- TD의 **Fluentd** 를 활용하여 **2시간 만에** 이용자 행동패턴 핵심 데이터 수집 기반 구축
- 모든 고객 상호작용 포인트마다 A/B 테스트를 실행하여 결과가 자동 분석되고 최적 대안도 자동으로 서비스에 적용되는 프레임워크 구축
- TD 웹콘솔과 Python API 를 이용하여 전직원 누구나가 필요한 데이터에 마음대로 접근해서 간단한 SQL 쿼리로 필요한 분석을 실시간으로 마음대로 수행하는 환경 제공
- 상품 판매자들에게도 개인화된 판매 분석 대시보드 제공하여 효과적인 상품 전략을 세울 수 있게 함

## ❖ 활용 결과

- 일 34억 이벤트 데이터를 TD로 처리하여 서비스/비즈니스 밸류 발굴
- 개인별로 최적화된 쇼핑 경험을 제공 → 고객이 WISH를 사랑하게 됨!
- 일 활성화 이용자 2억명 이상. 두번째로 많이 다운로드 된 쇼핑 앱
- 데이터 기반 자동화/효율화를 통한 멀티 버티컬 도입 및 성공





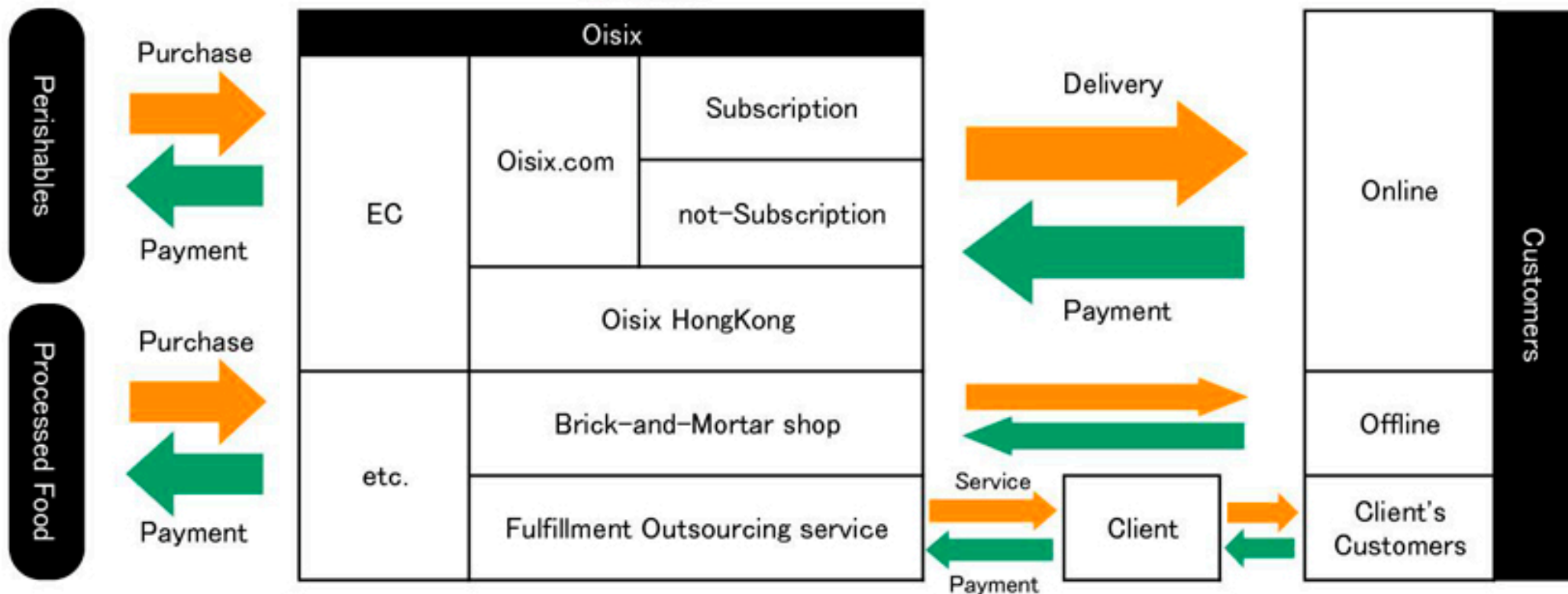
## 데이터로 돈 버는 글로벌 사례 #3

**ISISX : 머신러닝으로 이탈고객 예측&방지**



## Oisix - 회사 개요

- 2000년 창업. 자본금 약 84억원
- 식재료와 완제품 요리의 전자상거래/리테일 판매 비즈니스
- 가입회원 대상 지속적 판매를 하는 서브스크립션 커머스 중심





## Oisix – 회사 개요

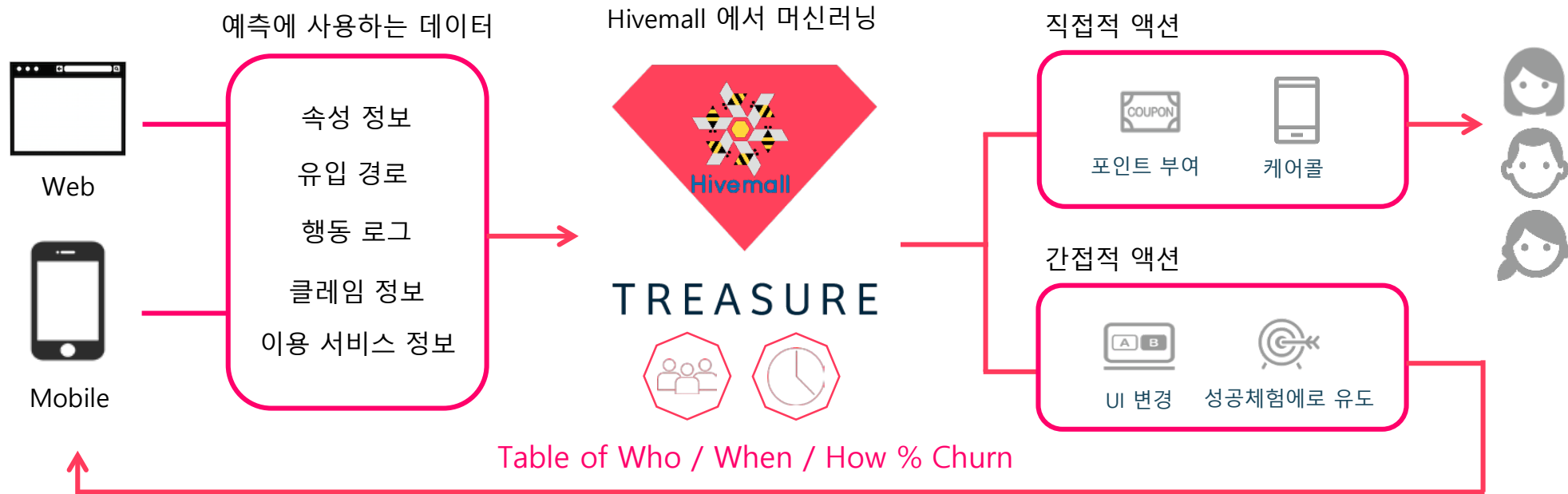
- 전통적인 서브스크립션 커머스 대비 회원비 부담 없고, 개인화된 구매(식재료의 종류&수량, 배달 시간)가 가능





# Oisix - 머신러닝을 활용한 이탈자 예측과 대응

## 고객 정보나 행동 로그를 바탕으로 기계학습을 하고 해지고객을 예측



<p><b>과제</b></p>	<ul style="list-style-type: none"> <li>회원의 정기 구매가 회사 전체 매출과 이익을 좌우하는데, 해지 징후가 있는 회원을 미리 파악 및 방어하는 대책이 부족했다</li> </ul>	<p><b>고객 사례</b></p>	<ul style="list-style-type: none"> <li>오이식스(Oisix)</li> </ul>
<p><b>실시 내용</b></p>	<ul style="list-style-type: none"> <li>기계학습을 하고 지난 1달간의 데이터를 바탕으로 앞으로 1달간에 해지 할 가능성이 높은 고객 리스트를 Hivemall을 이용하여 작성</li> </ul>	<p><b>기대 효과</b></p>	<ul style="list-style-type: none"> <li>통계의 전문 지식 없이 머신러닝</li> <li>해지 예측 리스트에 있는 회원에게 포인트를 부여함으로써 해지율 반감</li> <li>해지 원인이 되는 액션, 이벤트를 찾아내고 비해지고객의 특징적인 행동도 파악 가능</li> </ul>

# ◆ Oisix – 머신러닝을 활용한 이탈자 예측과 대응 w/ Hivemall

- Machine Learning made easy for SQL developers (ML for the rest of us)
- Interactive and Stable APIs w/ SQL abstraction



```
1 public final class TrainNewsGroups {
2
3     public static void main(String[] args) throws IOException {
4         File base = new File(args[0]);
5
6         Multiset<String> overallCounts = HashMultiset.create();
7
8         int leakType = 0;
9         if (args.length > 1) {
10             leakType = Integer.parseInt(args[1]);
11         }
12
13         Dictionary<String, Integer> dictionary = new Dictionary<>();
14
15         NewsGroupHelper helper = new NewsGroupHelper(dictionary);
16         AdaptiveLogisticRegressionWrapper wrapper = new AdaptiveLogisticRegressionWrapper(helper);
17
18         new AdaptiveLogisticRegressionWrapper(wrapper);
19         LearningAlgorithm learningAlgorithm = new LearningAlgorithm(wrapper);
20
21         List<File> files = new ArrayList<>();
22         for (File file : base.listFiles()) {
23             if (newsGroupHelper.isNewsGroup(file)) {
24                 files.add(file);
25             }
26         }
27
28         Vector v = helper.encodeFeatureVector(files, actual, leakType, overallCounts);
29         learningAlgorithm.train(actual, v);
30
31         Collection<LearningAlgorithm> learners = new ArrayList<>();
32         System.out.println("Starting to analyze state");
33         SGDInfo info = new SGDInfo(learningAlgorithm);
34         SGDHelper.analyzeState(info, leakType, k, best);
35         learningAlgorithm.close();
36         SGDHelper.dissect(leakType, newsGroups, learningAlgorithm, files, overallCounts);
37         System.out.println("Exiting main");
38     }
39 }
40
```

100+ lines of code

## Classification with Mahout

```
60 Collections.sort(counts, ordering.natural().reverse());
61 k = 0;
62 for (Integer count : counts) {
63     System.out.println(k + "\t" + count);
64     k++;
65     if (k > 1000) {
66         break;
67     }
68 }
69 }
70
```

```
CREATE TABLE lr_model AS
SELECT
    feature, -- reducers perform model averaging in parallel
    avg(weight) as weight
FROM (
    SELECT logress(features,label,..) as (feature,weight)
    FROM train
) t -- map-only task
GROUP BY feature; -- shuffled to reducers
```

This SQL query automatically runs in parallel on Hadoop





# 데이터로 돈 버는 글로벌 사례 #4

워너브로스 : 프로세스 자동화로 시간과 돈 절약

# WARNER BROS. ENTERTAINMENT

 adjust

swrve

App Annie

 UPSIGHT



## Pain Points

- 게임 서비스 사업을 진행하기 위해서는 다양한 솔루션의 도움이 필요. 각 솔루션별로 제공하는 레포트와 기능들을 활용
- 사업이 고도화될수록 각 솔루션들이 기본으로 제공하는 것들에 만족하지 못하게 되고, 각각의 raw data 를 직접 연결해서 활용하여 고차원의 분석을 하고 싶은 욕구가 강해짐
- 정형화된 레포트를 추출해서 보는 것만이 아니라 각 부서별로 그 때 그 때 필요한 분석을 하기 위해서는 데이터 엔지니어의 도움을 필요로 하다보니 병목이 점점 발생
- iOS, Android 매출 데이터의 병합 및 정제가 매끄럽지 못한 상황이 종종 발생

# WARNER BROS. ENTERTAINMENT



650B records stored!



- UDFs, Web-hooks, Scripts
- Interactive queries via Presto
- Significantly reduced latency



## Benefits by Treasure Data

- 활용하던 솔루션들의 raw data 를 트레저데이터로 자동으로 수집 통합하여 원하던 고차원 분석을 마음대로 할 수 있게 됨
- 데이터 엔지니어의 도움을 최소화하여 분석이 필요한 누구든지 인터랙티브 쿼리의 프레스토 기반으로 Adhoc 분석을 활용하게 됨
- 데이터 활용 프로세스에서의 병목이 대폭적으로 감소했을 뿐만 아니라, 허드렛 일을 하던 데이터 엔지니어 리소스의 비중을 줄이고 데이터 사이언티스트 비중을 훨씬 강화
- 트레저데이터가 만든 data workflow management engine 인 digdag.io 를 이용하여 매출 데이터 병합 및 정제와 이를 기반으로 한 매출 레포트 생성을 완전히 자동화



## Workflow

ID	42201
Name	nasdaq_analysis
Project	<a href="#">nasdaq_project</a>
Revision	fc62ee2b-8ed0-4637-a794-d514ef5ab641

## Definition

```

timezone: UTC

schedule:
  daily>: 07:00:00

_export:
  td:
    database: workflow_temp

+task1:
  td>: queries/daily_open.sql
  create_table: daily_open

+task2:
  td>: queries/monthly_open.sql
  create_table: monthly_open

```

## Sessions

ID	Project	Workflow	Revision	Session Time	Last Attempt	Last Attempt Duration	Status
487018	<a href="#">nasdaq_project</a>	<a href="#">nasdaq_analysis</a>	fc62ee2b-8ed0-4637-a794-d514ef5ab641	2016-10-12T00:00:00+00:00	21 hours ago	36s	✓ Success
486559	<a href="#">nasdaq_project</a>	<a href="#">nasdaq_analysis</a>	fc62ee2b-8ed0-4637-a794-d514ef5ab641	2016-10-11T00:00:00+00:00	2 days ago	34s	✓ Success

## Files

queries/daily\_open.sql

```

SELECT
  TD_DATE_TRUNC('day', time), AVG(open) AS daily_avg_open, AVG(close) AS daily_avg_close
FROM
  sample_datasets.nasdaq
WHERE
  TD_TIME_RANGE(time, '2013-01-01', '2015-01-01')
GROUP BY
  1

```

queries/monthly\_open.sql

```

SELECT
  TD_DATE_TRUNC('month', time),
  AVG(daily_avg_open) AS monthly_avg_open,
  AVG(daily_avg_close) AS month_avg_close
FROM
  daily_open
WHERE
  TD_TIME_RANGE(time, '2013-01-01', '2015-01-01')
GROUP BY
  1

```



# 데이터로 돈 버는 글로벌 사례 #5

DENTSU 등의 애드테크(ADTECH) 회사들

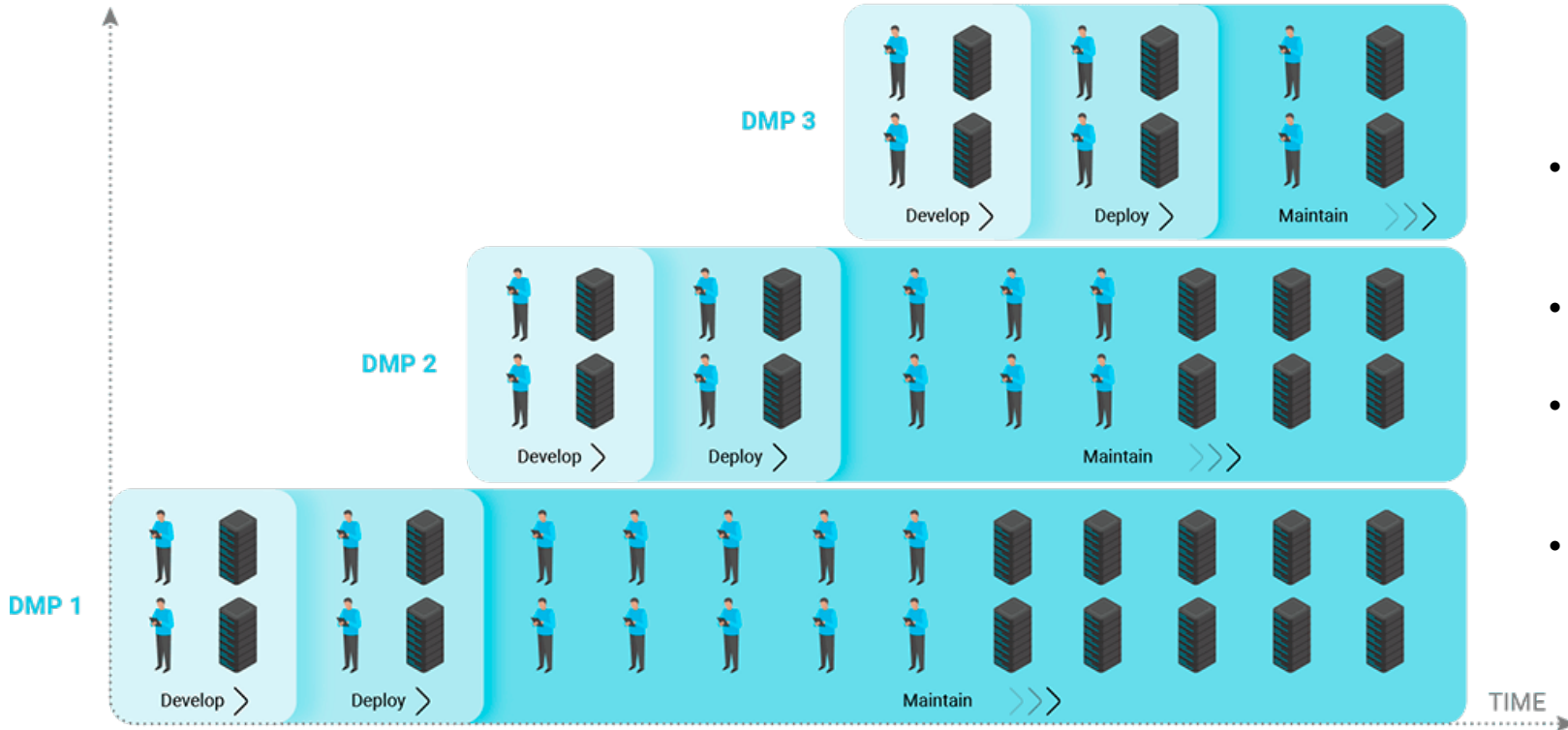


# Dentsu – Pain Points

Data Owner:  
**Dentsu**



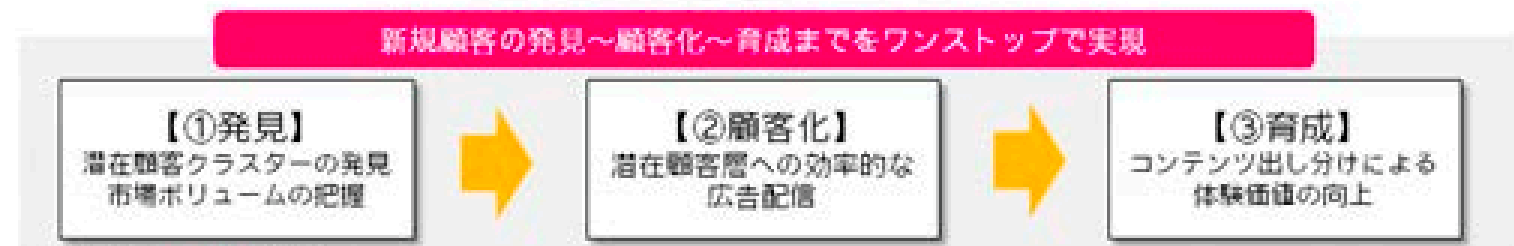
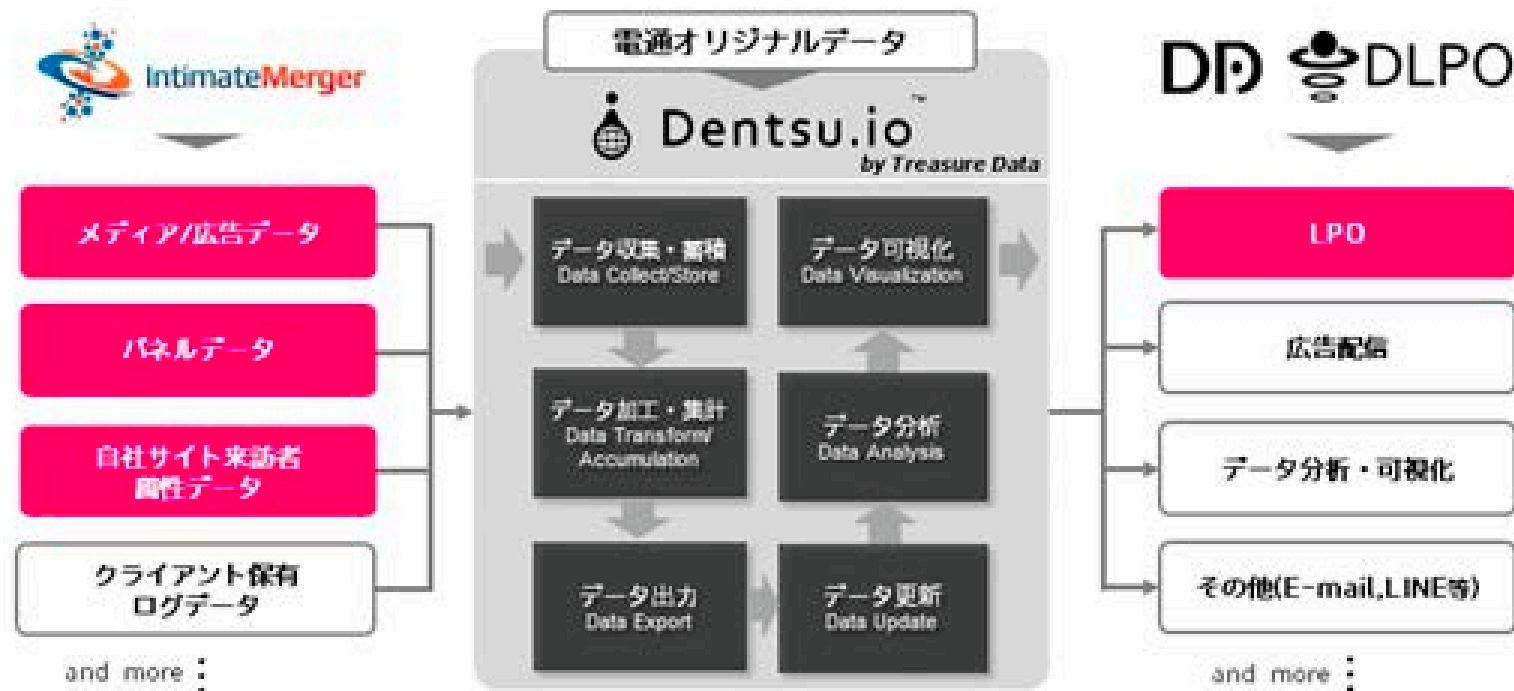
*In-house development of custom DMPs is resource-intensive and slow. And over time, more and more resources must be committed to maintenance.*



## Pain Points

- 자체 개발한 온프레미스 DMP 솔루션은 새로운 고객마다 데이터 소스를 처리하는 커스터마이징을 하기 위해 수개월 이상이 걸림
- 데이터 엔지니어링 작업의 완료를 대기하느라 데이터 사이언스 팀의 업무 처리 속도가 저하됨
- **DMP** 솔루션에 대한 막대한 수요를 감당하지 못함
- 인프라스트럭처를 디자인하고 유지하는 것은 덴츠의 핵심 역량이 아님
- 데이터가 물리적으로 덴츠의 소유이다보니 데이터의 소유를 원하는 고객사 니즈 못맞춤

# ◆ Dentsu – DMP Infra by Treasure Data and 3<sup>rd</sup> Parties



※機能拡張により■部分が可能になりました。  
※図内の「メディア/広告データ」「自社サイトデータ」は、インティメート・マージャー保有の独自データを活用して生成されたものを指します。



# Dentsu – Benefits by Treasure Data

Data Owner:  
**Client**



*Treasure Data frees up resources and accelerates deployment of custom DMPs. And with self-service analytics, clients can make changes on the fly.*



## Benefits by Treasure Data

- 별다른 추가 개발 없이 기존의 온프레미스 레거시 시스템에 바로 연동되는 유연한 클라우드 인프라를 통해 즉각적으로 문제가 해결됨
- 플러그인플러그 방식의 데이터 커넥터를 활용하여 새로운 데이터 소스를 설정하고 처리하는 데에 걸리던 시간을 수개월에서 몇일로 단축
- 엔지니어링 팀의 지원 없이 데이터사이언스 팀 스스로 새로운 디지털광고제품을 만들어냄
- 고객사의 데이터는 고객사 소유로, 고객사의 프라이빗 클라우드에 존재하여 마음대로 활용 가능



# Treasure DMP 를 활용한 디지털 마케팅



### 1ST Party (고객 데이터)

- 광고 로그 데이터
- 웹로그 데이터
- 모바일앱 데이터
- 소셜 데이터
- POS 데이터
- CRM 데이터
- 마케팅 자동화 데이터
- 매장 방문 데이터
- 고객 속성 데이터

## TREASURE DMP

원천빅데이터 저장, Data (ID) 통합  
세그멘테이션 / 대시보드  
데이터웨어하우스, 머신러닝 (추천 엔진 등)

**TREASURE COLLECTOR**  
실시간으로 다양한 빅데이터를  
간편하게 수집

**TREASURE RESULT**  
다양한 프로그램과  
쉽게 연동

**TREASURE EXCHANGE**  
기업이 보유하고 있지 않은  
이용자 ID의 연동

### 2nd, 3rd Party DMP

연계 데이터	날씨
추정 연간 소득	생애 주기
회사간 연결 데이터	관심사
POS 데이터	위치 데이터

3rd Party Cookie 연계 포함

### 광고 활동

- Ad Network
- DSP

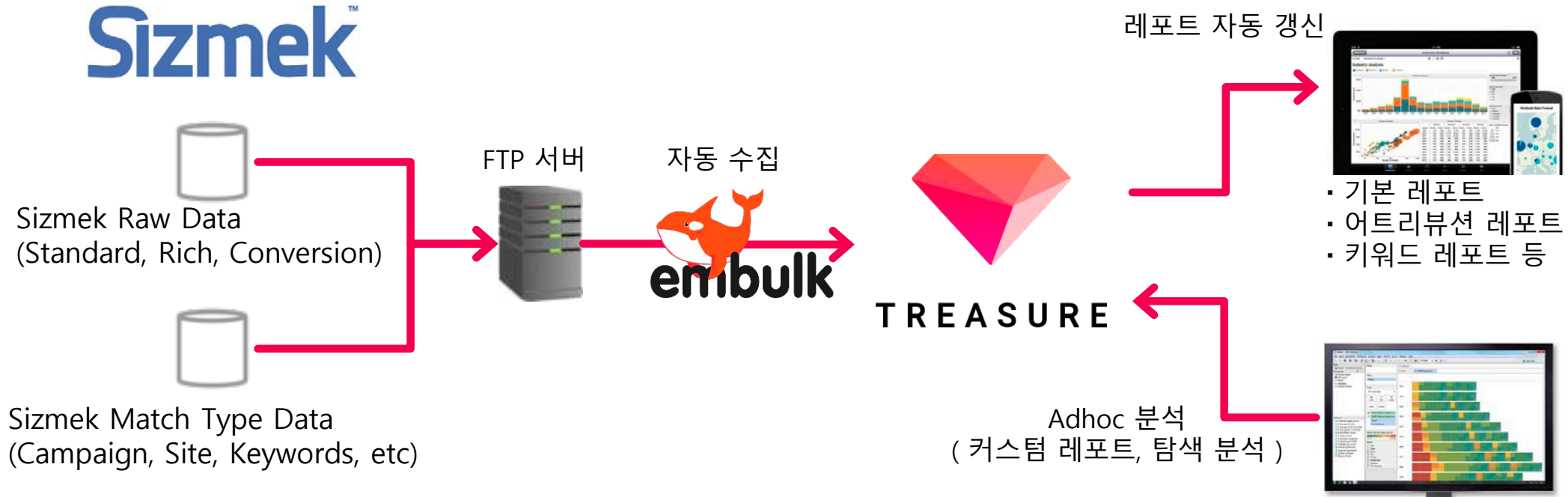
### 마케팅 활동

- Facebook / Twitter
- LINE
- Push Notification
- Divided out PC / SP site

### CRM 활동

- 이메일
- 콜센터

# ◆ 흩어져 있는 파일 형태의 광고 데이터 통합, 분석, 시각화



- ☹️
- 매번 정형 보고서를 만드는 데에 단조로운 작업이 많이 필요
  - 보고 싶은 보고서를 보는 데에 많은 화면 전환이 필요
  - 복수의 데이터 테이블을 다운로드하고 엑셀로 재집계하는 작업이 필요
  - Web UI 에서 조작하기에는 불편한 점들이 있음



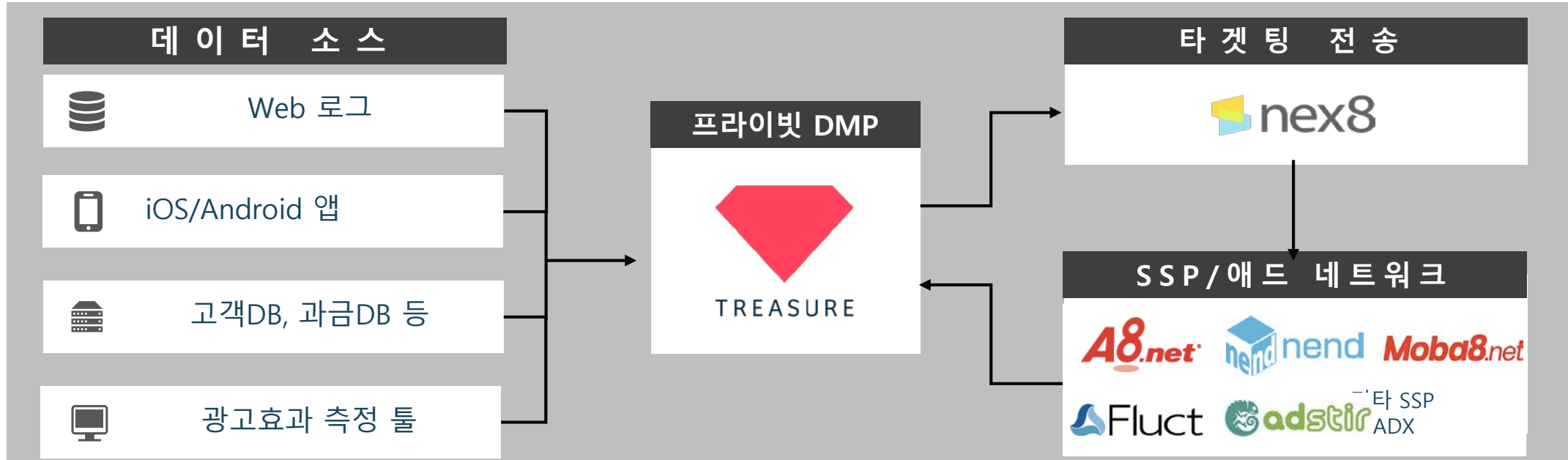
- 매일 CSV로 만들어지는 Sizmek 전체 원천 데이터를 트레저데이터로 자동 수집 저장
- 1차로 정제한 후에 Tableau 서버에 자동으로 전송
- Tableau Server 에서 기본 레포팅을 하고, Tableau Desktop 에서 Adhoc 분석을 실행



- 정형 보고서를 자동화할 수 있게 됨
- 보고 싶은 레포트를 Tableau 에서 손쉽게 볼 수 있게 됨
- 빅데이터의 고속 집계 및 애드혹 분석이 가능하게 됨
- 원천 데이터를 보존하고 있기 때문에, 모든 분석 요소 축과 조건을 활용한 시각화가 가능해짐

# ◆ 리인게이저먼트 광고 실행

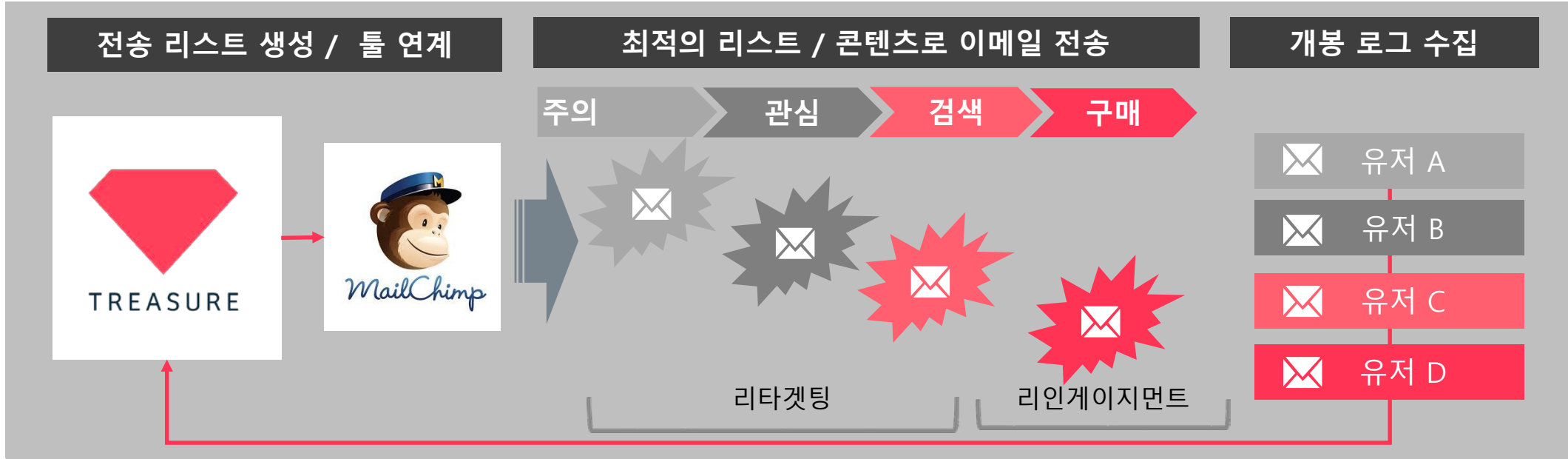
축적하고있는 세그먼트 데이터를 NEX8와 연계하여 타겟팅 전송



<p><b>과제</b></p>	<ul style="list-style-type: none"> <li>가망고객은 다 신규고객으로 만들어 놓은 상황이라 휴면 유저 발굴을 하고 싶다 (리텐션)</li> <li>PDCA사이클을 자동화 하고 싶다</li> <li>IDFA를 안전하게 연계시키고 싶다</li> </ul>	<p><b>적합 고객</b></p>	<ul style="list-style-type: none"> <li>모바일 앱 서비스 사업 회사</li> </ul>
<p><b>실시 내용</b></p>	<ul style="list-style-type: none"> <li>TD에서 작성한 세그먼트를 바탕으로 Nex8와 연계하여 광고를 전송한다.</li> <li>전송 결과도 IDFA단위로 TD상에서 추적.</li> <li>그 결과를 바탕으로 재분석하고 효과를 높인다.</li> </ul>	<p><b>기대 효과</b></p>	<ul style="list-style-type: none"> <li>IDFA를 기반으로 한 광고 전송</li> <li>휴면유저 발굴</li> <li>리셋 노가다 유저에 대한 광고 전송 제외</li> <li>PDCA고속 사이클에 의한 전송 최적화</li> </ul>

# ◆ 고도화된 타겟팅 이메일 자동화

## 고수준의 전송 리스트를 이용한 이메일 마케팅의 실현

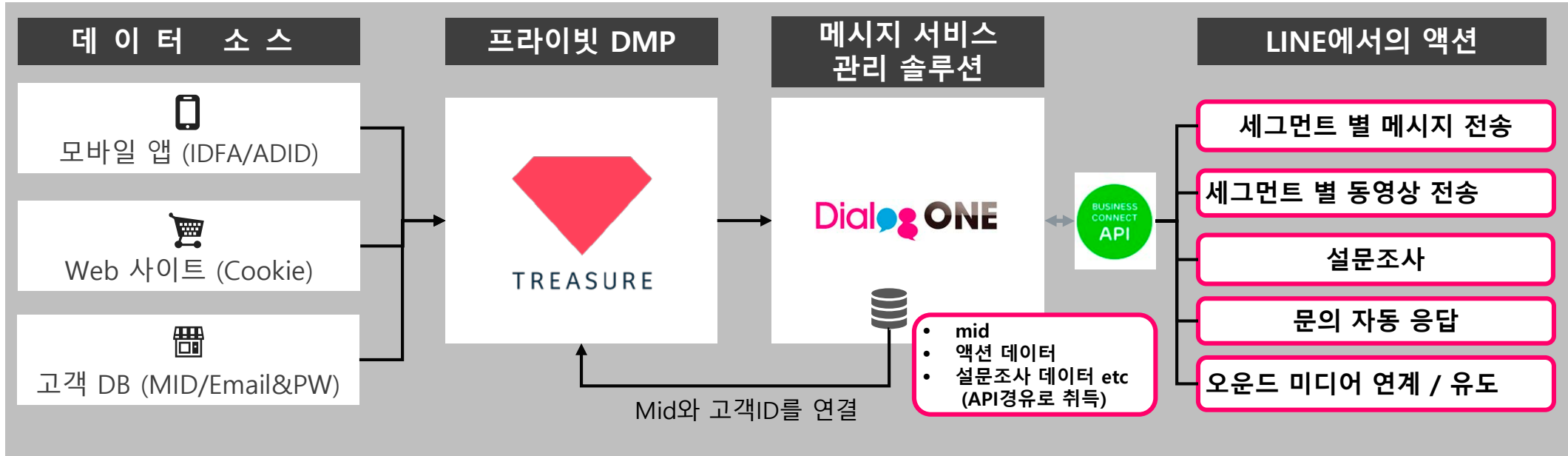


<p><b>과제</b></p>	<ul style="list-style-type: none"> <li>고객 여정의 단계와 속성에 맞게 최적화 된 메시지를 전송하고 싶다</li> </ul>	<p><b>적합 고객</b></p>	<ul style="list-style-type: none"> <li>회원 서비스 제공 기업</li> </ul>
<p><b>실시 내용</b></p>	<ul style="list-style-type: none"> <li>트레저데이터가 리스트를 작성 하고 이메일 전송 톨과 연계</li> <li>이메일 개봉 로그에서 전송 리스트를 자동 업데이트</li> </ul>	<p><b>기대 효과</b></p>	<ul style="list-style-type: none"> <li>고객여정의 단계, 고객 속성에 맞는 최적의 이메일 전송</li> <li>자동 처리에 의한 PDCA 고속화</li> </ul>



# LINE 메신저 플랫폼과 데이터를 활용한 광고

고객 정보와 연결시켜 유저마다 다른 콘텐츠를 LINE으로 전송



<p><b>과제</b></p>	<ul style="list-style-type: none"> <li>◆ 쌍방향 커뮤니케이션이나 1to1 마케팅 등 고객과의 관계를 강화 시키고 싶다</li> <li>◆ LINE을 사용한 새로운 액션을 실행하고 싶다</li> </ul>	<p><b>적합 고객</b></p>	<ul style="list-style-type: none"> <li>◆ B2C 기업</li> </ul>
<p><b>실시 내용</b></p>	<ul style="list-style-type: none"> <li>◆ LINE비즈니스 커넥트에서 취득 할 수 있는 유저 정보와 자사 보유 데이터를 연결시켜 콘텐츠나 광고를 전송</li> <li>◆ 그 결과를 바탕으로 재분석하고 효과를 높인다</li> </ul>	<p><b>기대 효과</b></p>	<ul style="list-style-type: none"> <li>◆ 유저의 로열티 상승</li> <li>◆ 정확도가 높은 광고 전송</li> </ul>



데이터로 돈을 벌고자 할 때  
반드시 체크해야 하는 것



## 트레저데이터가 주도하는 데이터 기반 혁신의 핵심 명제

- 분석을 위한 데이터가 아닌 실행을 위한 데이터. **Data Action**
- 데이터 기반의 프로세스 자동화 → **Data Workflow Management**
- 이종의 원천 데이터 결합을 통한 혁신 : **Linked Data Based Innovation**
- 데이터 기반의 **커뮤니케이션과 문화**
- Time based, streaming, log, ... : 인생은 타이밍, **데이터도 타이밍**
- **Data Product !!**
- **Data Scientist vs. Data Engineer ?!**
- Solution? Should be **EASY !!**

'트레저데이터'에 대해  
더 알고 싶으시면 클릭~!!



**고영혁**

트레저데이터 데이터사이언티스트/아키텍트  
dylan@treasure-data.com