M. Cloud Bridge
Specialized Consulting Firm in Data & AI

Specialized Consulting Firm in Data & AI Cloud System

Data & BI
Big Data & AI
Data Flow & Automation
Data Infra & Security

빅쿼리를 사용한
고객 데이터 플랫폼(CDP) 구축

데이터에 가치를 더하여 고객의 성장에 공헌합니다.
Specialized Consulting Firm in **Data & AI** Cloud System

M. Cloud Bridge
Specialized Consulting Firm in Data & AI

Agenda

1. Customer Data Platform 주요기능
2. 데이터 통합 시연

Customer Data Platform 주요기능

Customer Data Platform은 고객의 여정에서 발생하는 다양한 데이터를 수집, 정제, 비교하여 고객 통합 프로파일을 생성합니다. 이를 통해 고객을 분류, 그룹화하여 예측 및 의사결정을 지원하며 도출한 통찰력을 다양한 채널을 통해 활성화하는 핵심 5단계로 구현될 수 있습니다.

Customer Data Platform 핵심 5단계

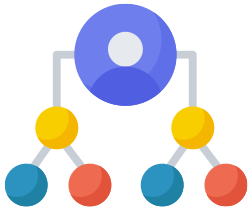
Data Collection

- 모든 소스에서 데이터 수집
- 사용 가능한 형태로 저장

Profile Unification

- 데이터 정제
- 중복 데이터 제거
- 사용자와 ID를 연결

Segmentation



- 고객 분류 및 그룹화

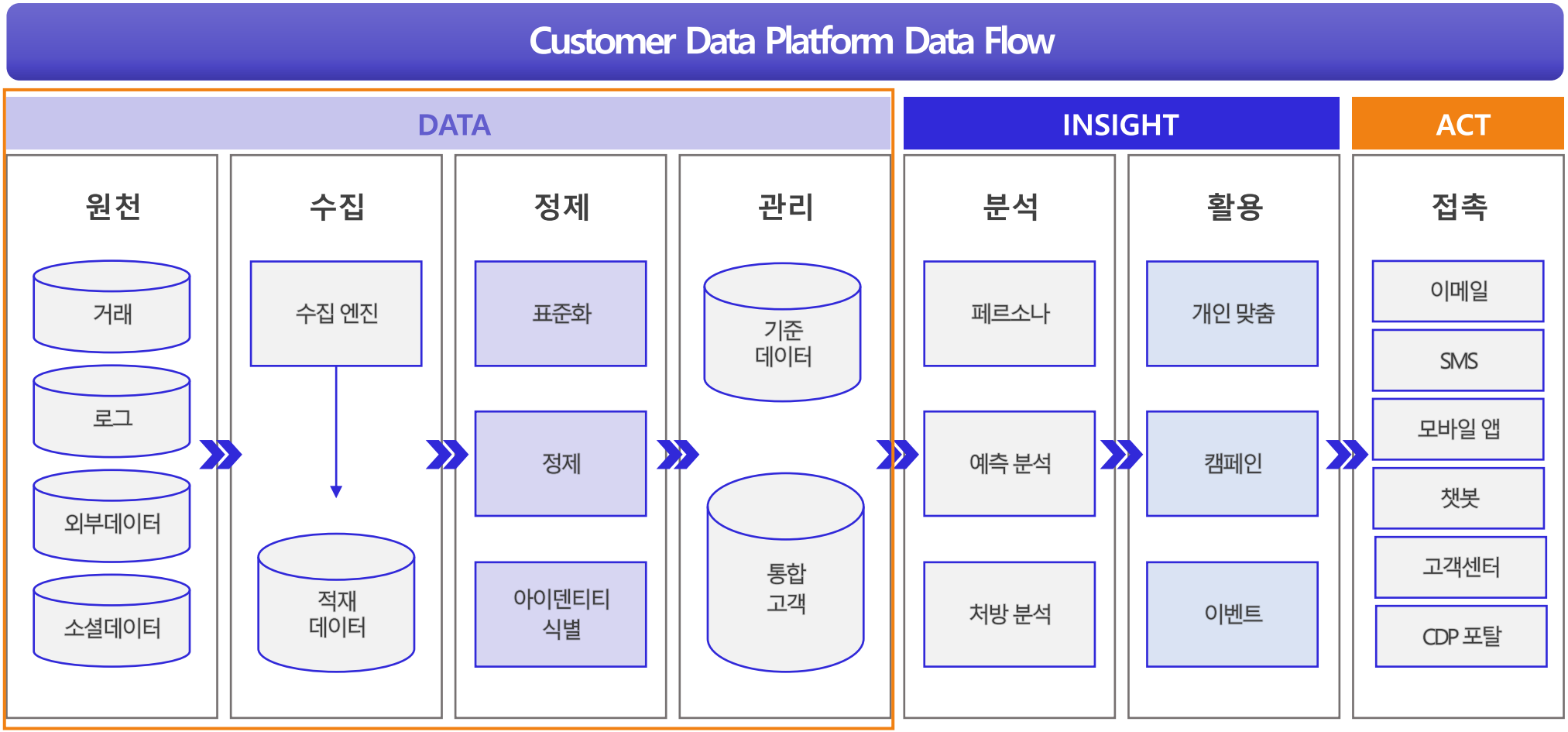
Prediction & Decision

- 예측 분석
- 의사결정 추천
- 개인 맞춤

Activation

- 이메일
- SMS
- 모바일 앱
- SNS
- 우편발송
- 챗봇
- 고객센터
- CDP 포탈

Customer Data Platform의 단계는 Data Flow의 관점에서 원천 데이터 선별, 데이터 연결 및 수집, 데이터 정제(표준화/정제/ID 식별), 통합고객 데이터 관리, 다각적인 분석, 활용 전략 도출, 효율적인 채널 연결 7가지의 상세 프로세스로 세분화 할 수 있습니다.



Customer Data Platform Data Flow에서 분석 이전 단계의 가장 중요한 부분은 고객 데이터 통합 절차라고 할 수 있으며, 각각의 데이터 원본이 수집되면 표준화, 정제등 전처리 작업을 거쳐 ID 식별, 고객 데이터 병합(Merge) 작업을 통해 Customer Single View(360 View)구축을 목표로 합니다.

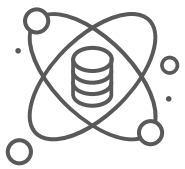
데이터 통합 절차

데이터 원본	데이터 수집	표준화	정제	매칭	데이터 통합
					
• 정형 • 비정형(반정형) • On/Offline	• GCS • BigQuery • Dataflow • Datafusion • Dataprep • DataFunction • Cloud Composer	• 데이터 포맷 • 데이터 분할 • 형식 통일 • 명칭 통일	• Null처리 • 중복제거 • 이상치 처리	• 기준 ID 매칭 • 정보 매칭 - Fuzzy - Deterministic - Probabilistic	• Main Table • History Table • Single View (360 View)

데이터 통합 시연

데이터 원본 선택은 다양한 고객의 여정을 이해하기 위해 온라인과 오프라인 및 각각의 접점(Touch Point 또는 Data Point)에서 발생하고 수집이 가능한 정형, 비정형, 반정형 형태의 데이터를 품질, 양, 비용 등을 고려하여 Customer Data Platform 구축 목적에 맞도록 선정하는 단계입니다.

데이터 통합 절차 - 1 데이터 원본 선택

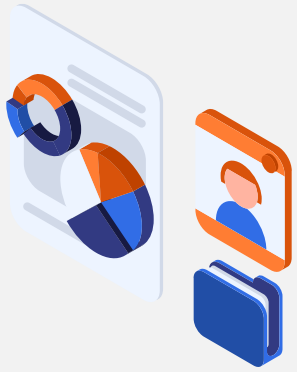
데이터 원본	데이터 수집	표준화	정제	매칭	데이터 통합
					
• 정형 • 비정형(반정형) • On / Offline	• GCS • BigQuery • Dataflow • Datafusion • Dataprep • DataFunction • Cloud Composer	• 데이터 포맷 • 데이터 분할 • 형식 통일 • 명칭 통일	• Null처리 • 중복제거 • 이상치 처리	• 기준 ID 매칭 • 정보 매칭 - Fuzzy - Deterministic - Probabilistic	• Main Table • History Table • Single View (360 View)

Customer Data Platform 구축에 필요한 데이터 원본은 고객의 여정에서 집계된 내부 데이터, 여정에 영향을 주는 기준 데이터, 외부에서 구입한 데이터, SNS 데이터 등 고객을 더 잘 이해하고 알맞은 서비스를 제공하는데 필요한 데이터들이 있으며 서비스 형태와 목적에 따라 다양합니다.

데이터 원본

고객데이터

1. 인구통계 데이터 (DB)
2. 고객 행동 데이터 (GA)
3. 사용 디바이스 이력 (GA)
4. 거래내역 (카트, 위시, 구매, 결재, 배송)
5. 캠페인 참여이력 (상품평, 설문조사 – Paper, CSV)
6. 고객 서비스 이력 (상담, A/S)
7. SNS 참여이력 (크롤링)
8. CRM 데이터 (DB)
9. 검색이력 (DB)
10. 마케팅 자동화 이력, 채널(반응) 로그 (CSV, API)
11. 오프라인 매장 방문 데이터 (Paper, CSV)
12. 판매 지점 데이터 (DB)
13. 할인 및 프로모션 데이터 (DB)

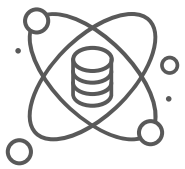


기준데이터

1. 유가
2. 시즌
3. 날씨(강우량, 강설량, 평균기온, 미세먼지)
4. 실업률
5. 기타 경제지표
6. 소비지표

Google Cloud Platform은 데이터 수집을 위해 Python, Java, Go, Scala, R, SQL 등의 대표적인 프로그램 언어를 지원합니다. 작업에 사용되는 Google 의 대표적인 서비스는 Cloud Storage, BigQuery, PubSub, Dataflow, Datafusion, Dataprep, Dataproc, Data Function, Cloud Composer 등이 있습니다.

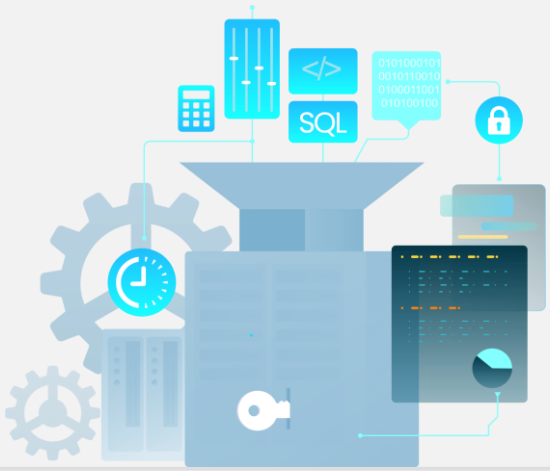
데이터 통합 절차 - 2 데이터 수집

데이터 원본	데이터 수집	표준화	정제	매칭	데이터 통합
					
• 정형 • 비정형(반정형) • On/Offline	• GCS • BigQuery • Dataflow • Datafusion • Dataprep • DataFunction • Cloud Composer	• 데이터 포맷 • 데이터 분할 • 형식 통일 • 명칭 통일	• Null처리 • 중복제거 • 이상치 처리	• 기준 ID 매칭 • 정보 매칭 - Fuzzy - Deterministic - Probabilistic	• Main Table • History Table • Single View (360 View)

데이터 수집 방법은 기업(Legacy)의 시스템 환경과 구축에 필요한 원본 데이터의 종류, 데이터 발생 시기와 포맷, 수집 주기, 분석/활용 등에 따라 매우 다양하며 여러 서비스의 연계를 통해 구현하거나 서비스 별 특징을 파악하고 목적에 맞는 서비스를 선택하여 data flow를 설계해야 합니다.

Data Collection Methods

1. Datasource > GCS(CSV) > Dataflow > BigQuery
2. Datasource > GCS(CSV) > Cloud Function > BigQuery
3. Datasource > GCS > Data Funsion > BigQuery
4. Datasource > Cloud SQL(PSQL, MySql, MSSQL) > Data Funsion > Dataprep > BigQuery
5. Datasource > GCS(CSV) > Dataprep > Cloud Functions > BigQuery
6. Datasource > Compute Engine > Pub/Sub() > Dataflow > BigQuery
7. Datasource > GCS(CSV) > Dataproc(Python, java, go) > BigQuery
8. Datasource > Cloud SQL > database migration(cdc) > BigQuery
9. Datasource > GA > BigQuery
10. Datasource > Cloud Dataproc > GCS > BigQuery
11. Datasource > Cloud SQL > BigQuery



Customer Data Platform을 구축할 때 필수적인 사용자 행동 데이터 집계를 위해 유용하게 사용되는 도구가 Google Tag Manager 입니다. 로그인 된 사용자의 정보와 행동 정보를 GA 등 마케팅 분석 도구에 전달하기 위한 코드를 생성해주고 테스트 기능을 제공합니다.

Data Collection Methods – Google Tag Manager

Google
Tag Manager

Tag Details

Properties

Name	Value
Type	Google 애널리틱스: GA4 구성
Firing Status	Not fired
설정할 필드	[{name: "user_id", value: "alan"}]
이 구성이 로드될 때 페이지 조회 이벤트 전송	true
서버 컨테이너로 전송	false
측정 ID	"G-

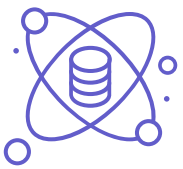
```
<!doctype html>
<html lang="ko">
<head>

<!-- dataLayer 변수 추가-->
<script>
  window.dataLayer = window.dataLayer || [];
  window.dataLayer.push({
    'userId': 'alan' //고객 번호와 같은 개인정보가 아닌 고유한 식별 값을 입력합니다.
  });
</script>
<!-- End dataLayer 변수 추가-->

<!-- Google Tag Manager -->
<script>(function(w,d,s,l,i){w[l]=w[l]||[];w[l].push({'gtm.start':
new Date().getTime(),event:'gtm.js'});var f=d.getElementsByTagName(s)[0],
j=d.createElement(s),dl=l!='dataLayer'?'&l='+l:'';j.async=true;j.src=
'https://www.googletagmanager.com/gtm.js?id='+i+dl;f.parentNode.insertBefore(j,f);
})(window,document,'script','dataLayer','GTM          ');</script>
<!-- End Google Tag Manager -->
```

Customer Data Platform의 데이터 통합과정에서 가장 먼저 수행되는 전처리 작업은 표준화 작업이며 수집된 데이터의 포맷을 통일시키고 각각을 의미 단위로 분할하며 서로 다른 순서와 양식을 가진 데이터는 표준 형식으로 변경하고, 컬럼 명을 최종 집계될 의미 명칭으로 매핑하는 것입니다.

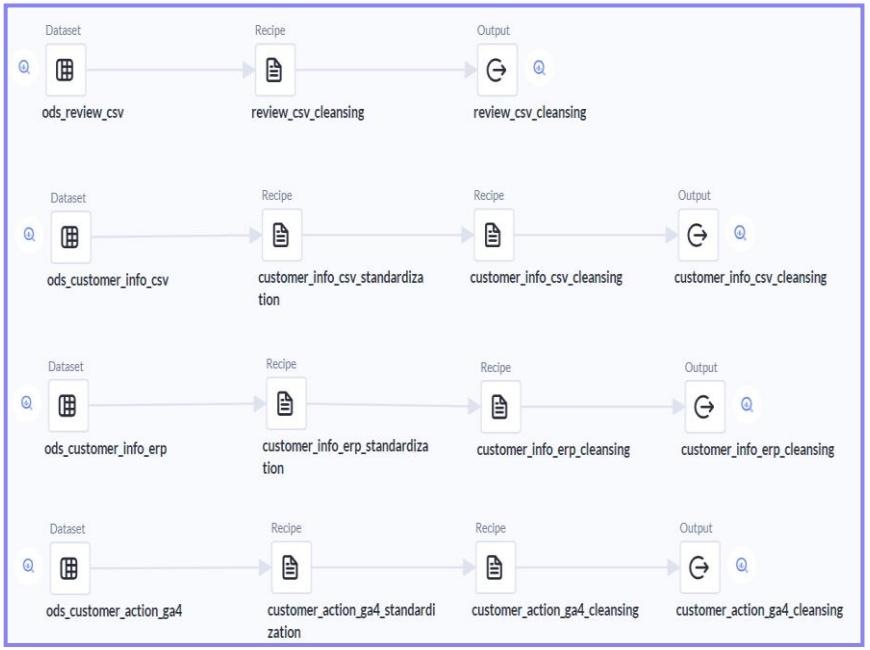
데이터 통합 절차 - 3 표준화

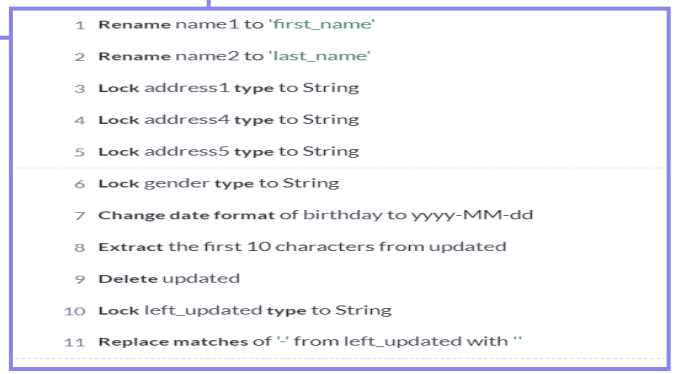
데이터 원본	데이터 수집	표준화	정제	매칭	데이터 통합
					
• 정형 • 비정형(반정형) • On / Offline	• GCS • BigQuery • Dataflow • Datafusion • Dataprep • DataFunction • Cloud Composer	• 데이터 포맷 • 데이터 분할 • 형식 통일 • 명칭 통일	• Null처리 • 중복제거 • 이상치 처리	• 기준 ID 매칭 • 정보 매칭 - Fuzzy - Deterministic - Probabilistic	• Main Table • History Table • Single View (360 View)

Dataprep을 활용한 표준화는 빅쿼리에 적재한 데이터를 Dataprep을 사용해 작업 후 다시 빅쿼리에 적재하는 과정을 수행합니다. Dataprep은 데이터를 시각적으로 탐색, 정리, 준비하는 지능형 클라우드 데이터 서비스이며 누락된 값을 식별하고, 중복을 제거하고, 데이터를 정규화하고, 변환할 수 있습니다.

데이터 표준화 - Dataprep

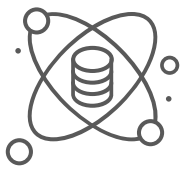
BigQuery » Dataprep » BigQuery





Customer Data Platform의 데이터 통합과정 중 데이터 매칭 작업 전 단계로 불필요한 작업 발생을 지양하고 오류 발생을 최소화 하기 위하여 데이터의 중복을 제거하고 기본값을 반영하거나 Null을 처리하며 구축 범위 밖 데이터를 필터링하고 생략된 용어나 누락된 부분을 추가하는 작업입니다.

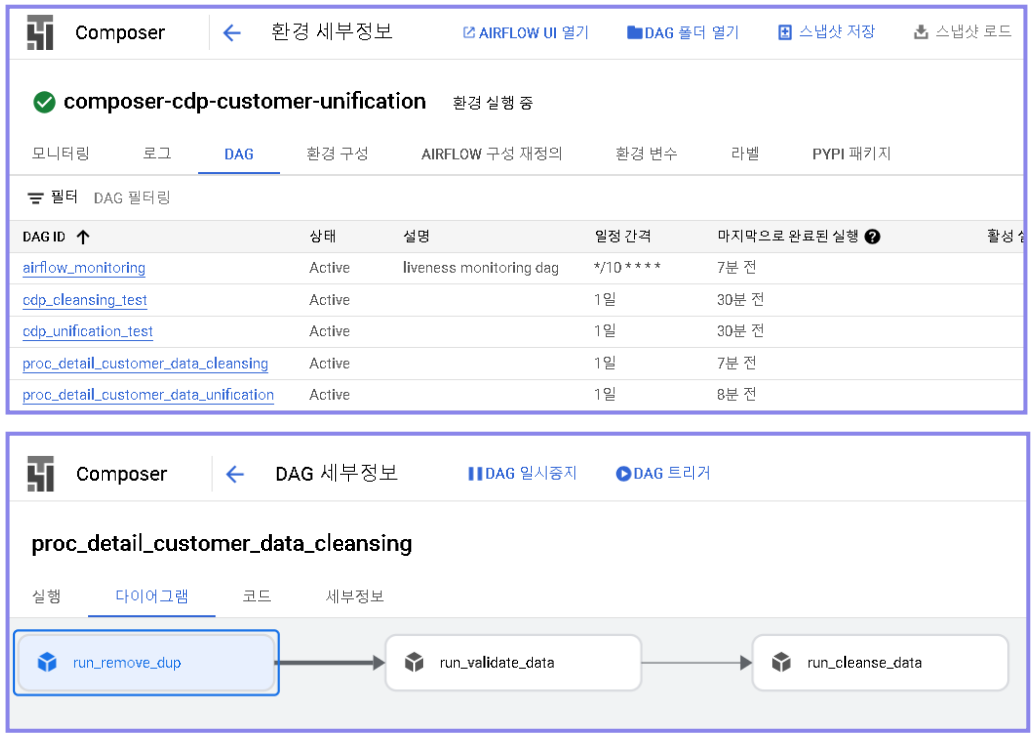
데이터 통합 절차 - 4 정제

데이터 원본	데이터 수집	표준화	정제	매칭	데이터 통합
					
• 정형 • 비정형(반정형) • On/Offline	• GCS • BigQuery • Dataflow • Datafusion • Dataprep • DataFunction • Cloud Composer	• 데이터 포맷 • 데이터 분할 • 형식 통일 • 명칭 통일	• Null처리 • 중복제거 • 이상치 처리	• 기준 ID 매칭 • 정보 매칭 - Fuzzy - Deterministic - Probabilistic	• Main Table • History Table • Single View (360 View)

Cloud Composer를 사용하여 대상 데이터를 정제하는 파이프라인을 구성할 수 있습니다. Google Cloud Composer는 Apache Airflow를 기반으로 하는 엔터프라이즈급 워크플로우 관리 서비스이며 데이터 수집, 데이터 처리, 데이터 분석, 데이터 시각화 등 다양한 작업을 자동화할 수 있습니다.

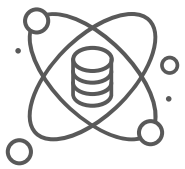
데이터 정제 – Google Cloud Composer

Google Cloud Composer



Customer Data Platform의 데이터 통합과정의 핵심이자 최종 단계인 매칭은 수집된 식별 정보들을 비교하여 신규 고객이나 기존 고객임을 식별하는 작업으로 다양한 매칭 방법론(퍼지, 결정론적 방법, 확률론적 방법 등)을 복합적으로 활용하여 정보를 매칭하고 머지하는 과정을 수행합니다.

데이터 통합 절차 - 5 매칭

데이터 원본	데이터 수집	표준화	정제	매칭	데이터 통합
					
• 정형 • 비정형(반정형) • On / Offline	• GCS • BigQuery • Dataflow • Datafusion • Dataprep • DataFunction • Cloud Composer	• 데이터 포맷 • 데이터 분할 • 형식 통일 • 명칭 통일	• Null처리 • 중복제거 • 이상치 처리	• 기준 ID 매칭 • 정보 매칭 - Fuzzy - Deterministic - Probabilistic	• Main Table • History Table • Single View (360 View)

Matching 방법론에는 다양한 기법들이 있으며 대표적으로 퍼지 방법론은 유사하지만 동일하지 않은 두 개 이상의 문자열을 비교하는 방법이며, 결정론적 방법은 두 개 이상 식별 정보의 100% 일치 여부를 확인하고, 확률론적 방법은 부수적인 식별 정보들을 통해 동일 고객일 것으로 추정하는 방법입니다.

Customer Identity Matching Method

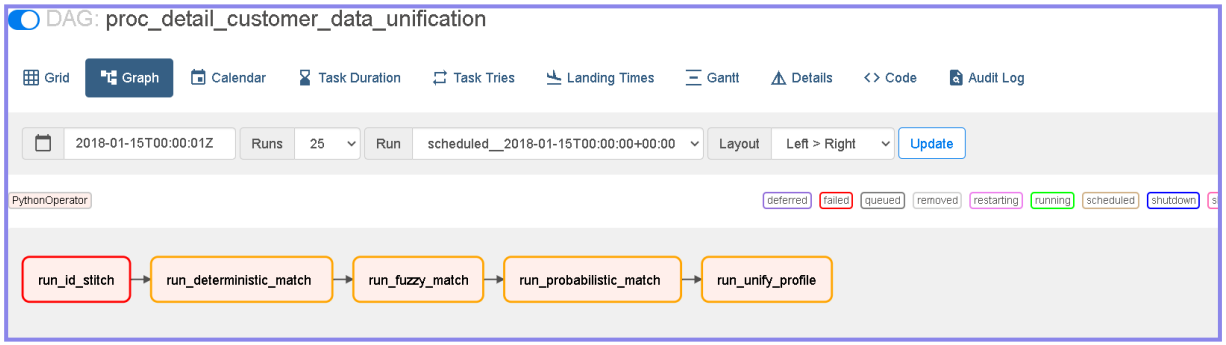
	Fuzzy Matching	Deteministic	Probabilistic
정의	유사하지만 동일하지 않은 두 개이상의 문자열 비교. 레벤슈타인 거리 또는 자로-윙클러 거리 등의 다양한 알고리즘을 사용	동일한 두 개 이상의 문자열을 매칭. 문자열을 한 글자씩 비교하는 등 다양한 방법을 사용	부수적인 식별 정보로 동일한 고객일 것이라는 통계적 가능성을 추정. 펠레기-선터 알고리즘, Jaccard 계수 등의 다양한 알고리즘을 사용
예시	• 이름, 집 주소 • Joe Smith = Joseph Smith	• 고객의 이메일 주소와 계정을 만들 때 사용한 이메일 주소 • mbinden@email.com = mbinden@email.com	IP 주소, 장치 유형, 운영 체제 또는 브라우저 유형 등을 사용함. 정적 일치보다 확실하지 않지만 범위를 확장하거나 제한된 자사 데이터를 보충하는 데 유용함.
사용 사례	• CRM • Customer support	• Email marketing • Social activation	• Scale and reach on the open web

Cloud Composer를 사용하여 대상 데이터를 통합 고객 ID와 매칭하고 하나의 마스터 테이블로 머지하는 파이프라인을 구성할 수 있습니다.

Google Cloud Composer의 Dag(Directed acyclic graph)를 활용하면 워크플로우를 시각화하여 관리하고 작업별로 모니터링하거나 자동화할 수 있습니다.

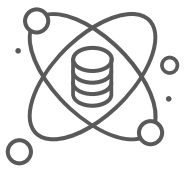
데이터 매칭 – Google Cloud Composer

Google Cloud Composer



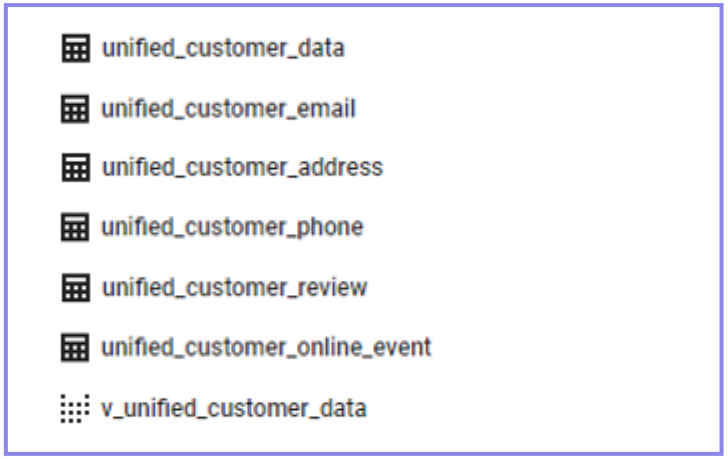
Customer Data Platform의 데이터 통합 절차의 최종 단계는 통합의 산출물이 생성되는 단계로 Customer Single View (또는 360 View) 가 생성되며 구축 목표에 따라 다양한 기준 데이터를 함께 관리하고 각종 보안 위협을 대비하여 DLP(Data Loss Prevention)기능을 통해 식별 정보를 보호합니다.

데이터 통합 절차 - 6 데이터 통합

데이터 원본	데이터 수집	표준화	정제	매칭	데이터 통합
					
• 정형 • 비정형(반정형) • On / Offline	• GCS • BigQuery • Dataflow • Datafusion • Dataprep • DataFunction • Cloud Composer	• 데이터 포맷 • 데이터 분할 • 형식 통일 • 명칭 통일	• Null처리 • 중복제거 • 이상치 처리	• 기준 ID 매칭 • 정보 매칭 - Fuzzy - Deterministic - Probabilistic	• Main Table • History Table • Single View (360 View)

통합 고객 데이터 생성은 영구적인 통합 ID(Unified Identity)에 1:1로 매칭되는 Main 데이터 테이블과 1:N의 형태로 제공되는 정보성, 이력성 테이블로 구성하며 다양한 채널을 통해 Single View 형태를 제공하기 위해 여러 테이블을 조인한 통합 View 테이블도 생성합니다.

고객 통합 데이터 생성




unified_customer_data

스키마 세부정보 미리보기 계보

행	uuid	custId	first_name	last_name	full_name	phone	email	address_detail
1	U000000322	C000322	Gerda	Basketter	Gerda Basketter	null	gbasketter9k@washingtonpo...	9 Mayfield Park
2	U000000956	C000956	Neddie	Beagin	Neddie Beagin	202-417-4920	nbeagingh@apple.com	6708 Del Sol Crossing
3	U000000968	C000968	Aaron	Peret	Aaron Peret	null	aperet59@hotmail.com	3233 Fuller Circle
4	U000001831	C001831	Mehetabel	Gowdy	Mehetabel Gowdy	213-911-7471	mgowdydg@qq.com	92649 Coleman Avenue
5	U000000440	C000440	Dillie	Messruther	Dillie Messruther	null	dmessrutherao@samsung.com	2912 Dayton Street
6	U000001084	C001084	Chariot	Jentle	Chariot Jentle	null	cjentleqg@buzzfeed.com	44 Union Point
7	U000000339	C000339	Conrade	Iannazzi	Conrade Iannazzi	202-654-5869	ciannazziex@cocolog-nifty.com	39688 Grover Trail
8	U000000391	C000391	Georgina	Tidman	Georgina Tidman	null	gtidmanhz@com.com	4620 Macpherson Point
9	U000000480	C000480	Sven	Ellerington	Sven Ellerington	null	selleringtonqu@freewebs.com	62730 Manley Terrace
10	U000000828	C000828	Saxe	Snell	Saxe Snell	null	ssnellkb@fastcompany.com	49126 Badeau Junction
11	U000000381	C000381	Georgiana	Grattage	Georgiana Grattage	null	ggrattageod@yahoo.com	332 Welch Crossing
12	U000000699	C000699	Vassily	Shaves	Vassily Shaves	312-832-8832	vshavesot@networksolutions...	8 Westend Court
13	U000001213	C001213	Sauveur	Zealy	Sauveur Zealy	null	szealyqc@forbes.com	8 Sloan Pass
14	U000001253	C001253	Tracie	Elvy	Tracie Elvy	814-838-8441	telvyq6@washington.edu	68883 Harper Alley
15	U000001600	C001600	Gayel	Sinyard	Gayel Sinyard	null	gsinyardfx@dot.gov	21744 Gateway Court
16	U000000291	C000291	Jourdain	Hallows	Jourdain Hallows	415-195-8356	jhallows6s@cisco.com	01525 Stang Avenue
17	U000000522	C000522	Carr	Rollitt	Carr Rollitt	239-268-2221	crollittkp@mashable.com	55658 Summerview Crossing
18	U000001507	C001507	Brigg	Follows	Brigg Follows	862-273-7951	bfollowsht@wufoo.com	108 Westridge Park
19	U000001914	C001914	Lorette	Crosser	Lorette Crosser	null	lcrosserj9@usnews.com	06 Evergreen Plaza

Customer Single View (또는 360 View)는 수집 집계된 1차 데이터 뿐만 아니라 이를 기반으로 고객을 분류하고 그룹화 한 데이터, 페르소나, 예측, 처방 분석의 내용들까지 포함하는 포괄적 데이터입니다. 각 분야별 담당자는 이 데이터에 접근하여 고객을 선별하고 마케팅 전략을 수립할 수 있습니다.

Customer Single View (360 View)

신선 식품 주구매 & 구매주기 5일 이내
→ 신선 정기 배송 안내

현재 Premium & APP Push 미동의
→ Push 동의(혜택 알리미) LMS 안내

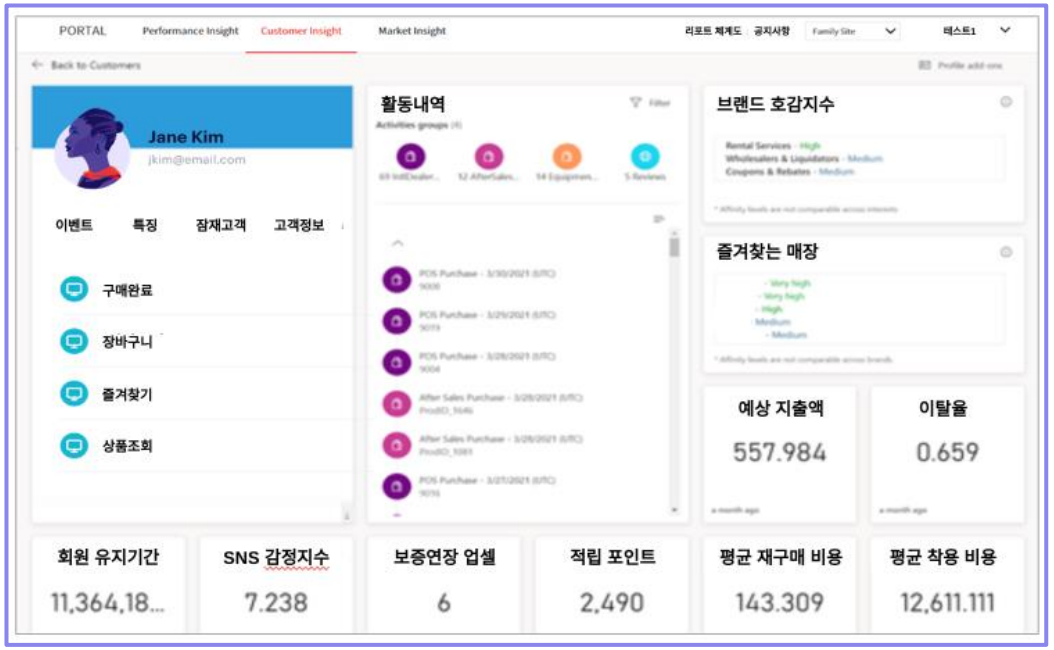
	Demo		쇼핑패턴						가격민감	채널선호	Value					Marketing Comm. & Response		
고객ID	성	연령	주구매 상품	구매주기 (일)	선호 시간대	주중/주말 선호	최근 방문 경과 일수	…	가격 민감도	온라인 채널가망	프리미엄 이탈예측	Value Segment	16주 구매금액 (만원)	16주 구매횟수	…	마케팅 활용	APP Push	…
UC0100XXX	여	50	신선	4	오후	주중	3	…	둔감	20	15	Premium	181	10	…	Y	N	…
UC0110XXX	여	40	비식품 그로서리	15	오전	주말	10		보통	60	-	Engaged	75	8		Y	Y	
UC0103XXX	여	34	식품 그로서리	28	저녁	주중	35	…	보통	85	-	Engaged	62	4	…	Y	Y	…
UC0201XXX	여	49	가정용품	50	오전	주중	62	…	보통	-	-	Swing	8	2	…	N	N	…
UC0340XXX	여	33	식품 그로서리	35	점심	주말	10		보통	45	30	Premium	112	3		N	Y	
UC0123XXX	여	42	신선	22	저녁	주말	15	…	민감	-	88	Premium	83	5	…	Y	Y	…

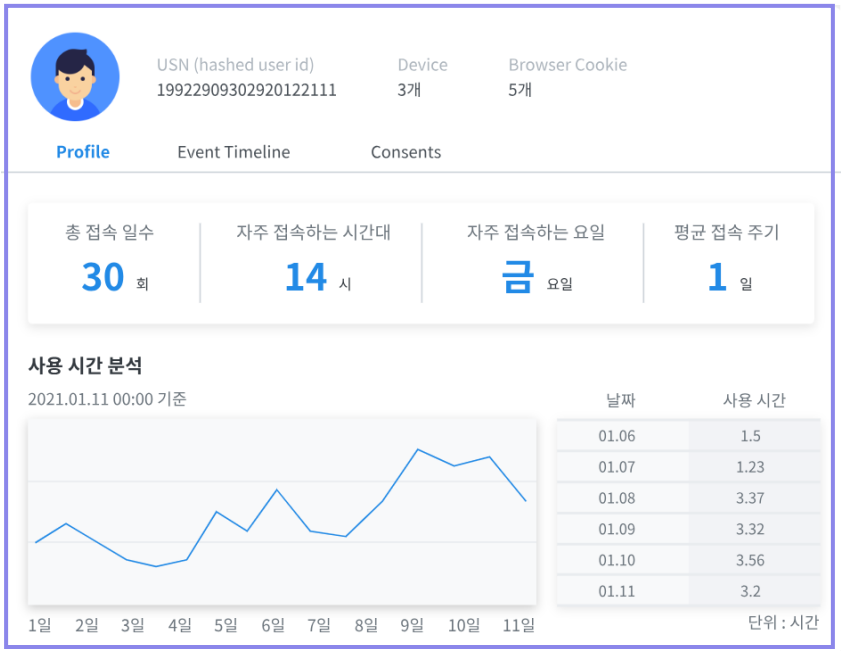
Only 오프라인 방문 & 온라인 가망 80점 이상
→ 오프라인 단골 상품으로 GHS 방문 유도

현재 Premium & 이탈예측 80점 이상
→ '한번 더 방문' 장바구니 쿠폰 발송

Customer Data Platform이 기존에 사용되어 오던 고객 관리 정보 시스템과 다른 점은 통합된 고객 데이터를 구축한다는 것과 이렇게 구축된 통합 데이터를 필요한 영역에 보다 쉽고 포괄적으로 제공한다는 것이며 대표적인 사례는 Customer Data Portal이나 Dashboard입니다.

Customer Data Portal & Dashboard





M. Cloud Bridge

Specialized Consulting Firm in Data & AI

Specialized Consulting Firm in Data & AI Cloud System

Data & BI

Big Data & AI

Data Flow & Automation

Data Infra & Security

Thank You

T. 02.552.9700

E. info@mcloudbridge.com

H. www.mcloudbridge.com

데이터에 가치를 더하여 고객의 성장에 공헌합니다.

Specialized Consulting Firm in **Data & AI** Cloud System