Cyber Security Conference 2021

SAMSUNG SDS

삼성SDS가 추구하는

프라이버시 강화 기술

한규형 프로 삼성SDS 보안연구센터

데이터 분석 및 활용 관련 규제 동향

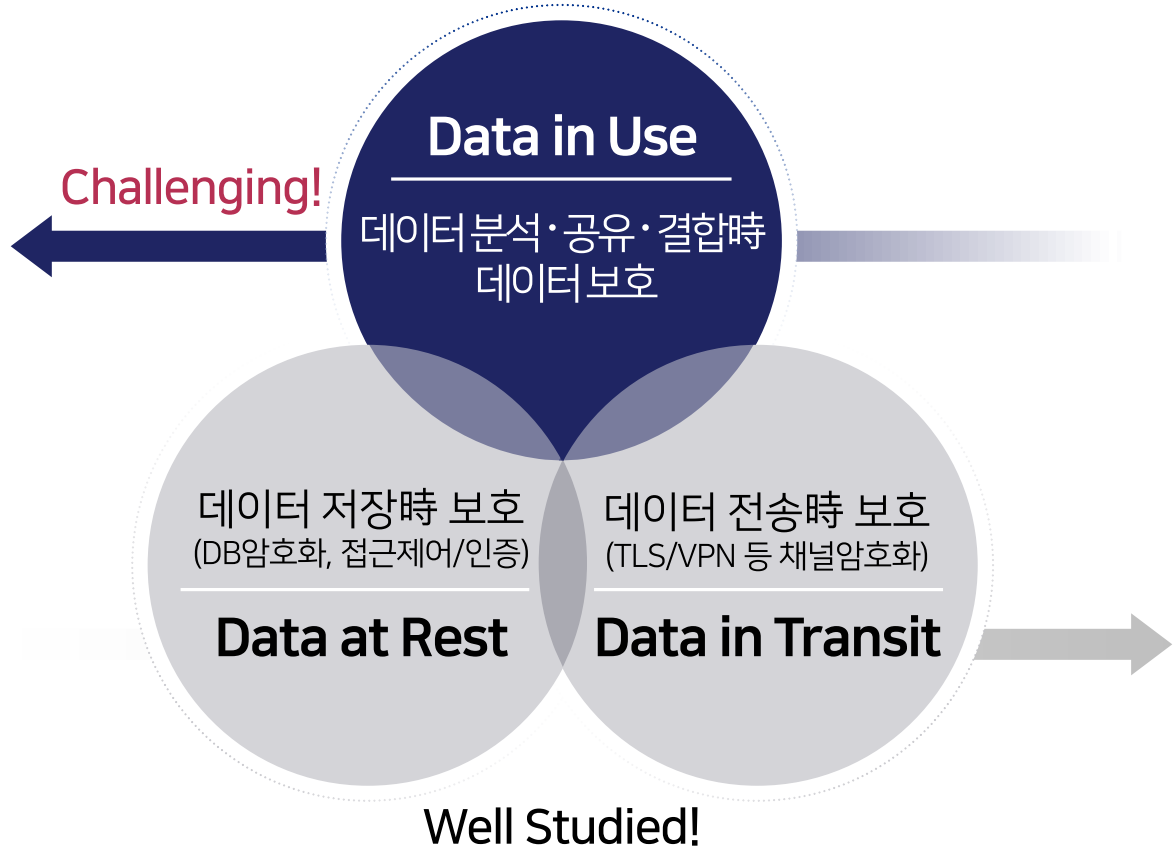
데이터 분석과 AI와 같은 기술의 활용 증가로 데이터 사용에 대한 필요성은 증가하고 있으나, 프라이버시 법규 강화로 데이터에 대한 안전한 활용 및 공유 방안 필요

❶ GDPR('18), CCPA('19) 이후 개인정보규제가 세계적으로 강화되고 있는 추세

- EU-US Privacy Shield 무효화 ('20.7月), 5000+ 기업이 유럽・미국간 데이터 전송時 이슈발생, 동형암호 등의 기술 사용필요 언급 (WSJ)

❷ 데이터의 안전한 공유를 통한 활용을 장려

- 국내 데이터3법을 통한 데이터 결합 가능 유럽보안정책기관(ENISA)에서의 데이터공유 時 PETs*기술 활용 장려



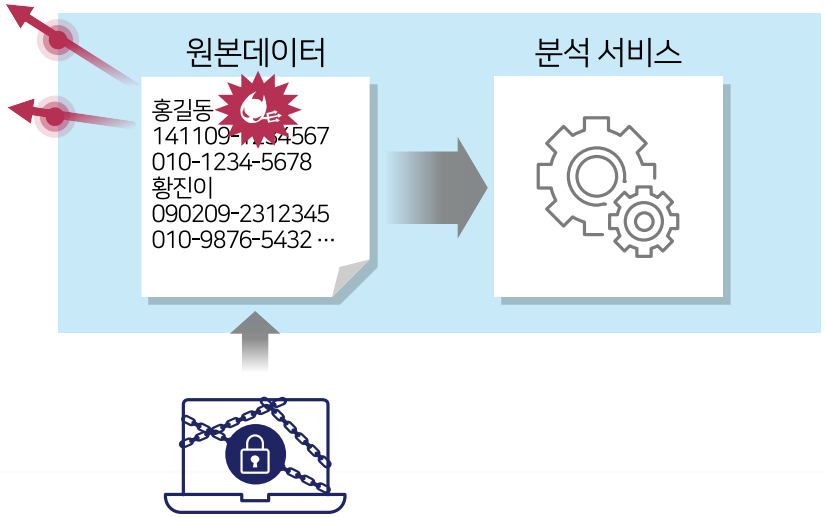
* Privacy Enhancing Technologies (PETs)

기존 기술의 한계점 – 표준 암호화 기술

데이터 처리 및 분석을 위해서는 비밀키 정보를 가지고 복호화를 수행해야 한다는 기술적 한계점

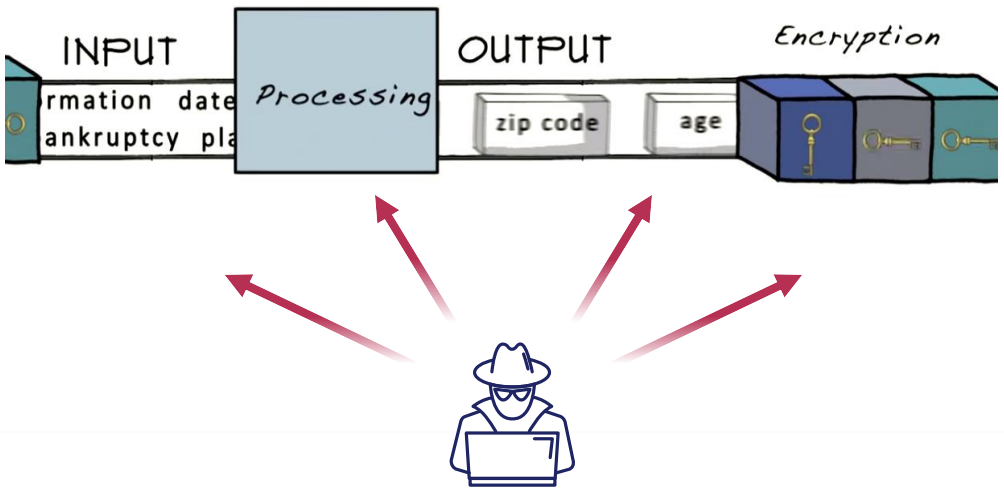
01

기존 표준 암호화 기술의 경우
데이터 처리를 위해서는 원본 데이터 필요



02

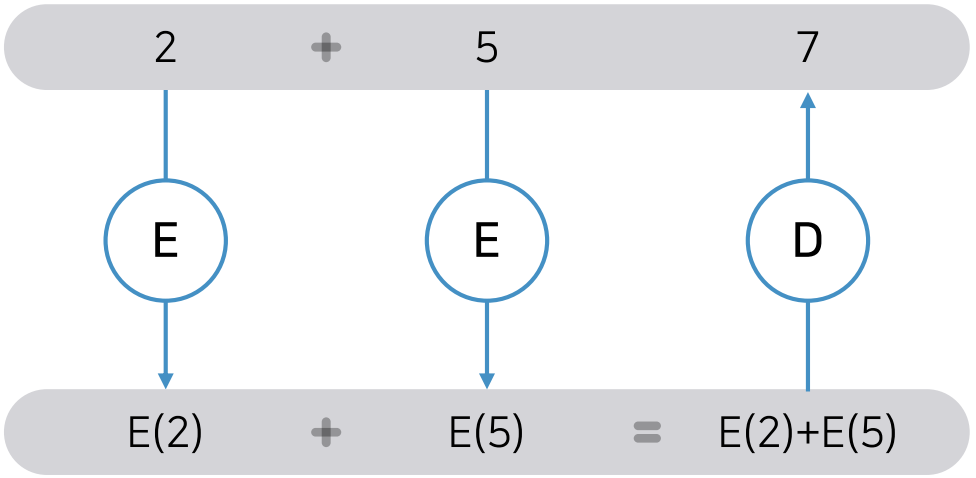
정보 처리 주체의 암호화 키에 대한 접근으로 인한
잠재적인 보안 위협 존재



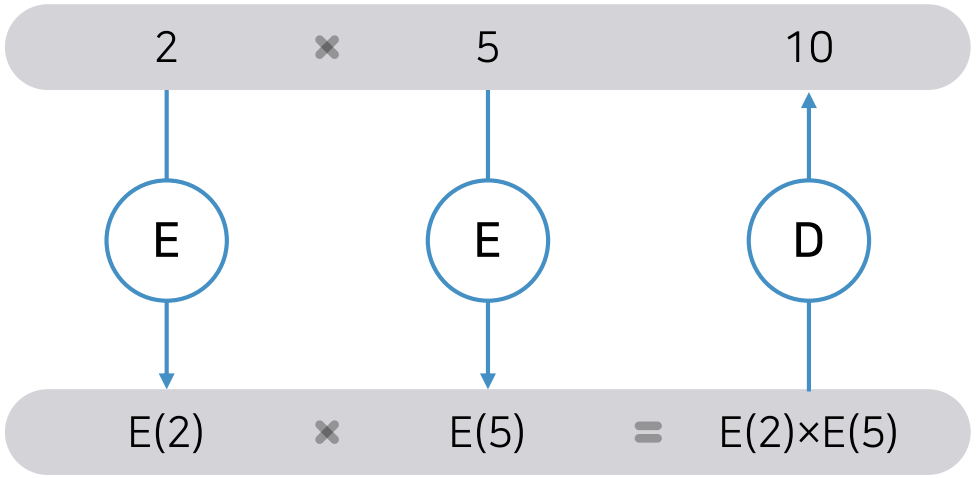
❶ 동형암호 (Homomorphic Encryption)

1978년 Rivest에 의해 고안된 나온 개념, 2009년 처음 Gentry에 의해 설계됨

덧셈 연산 보존



곱셈 연산 보존



“ 동형 암호는 덧셈/곱셈 연산을 보존하여,
암호화된 데이터를 활용한 데이터 분석이 가능하도록 지원하는 차세대 암호기술 ”

❶ 동형암호 (Homomorphic Encryption)

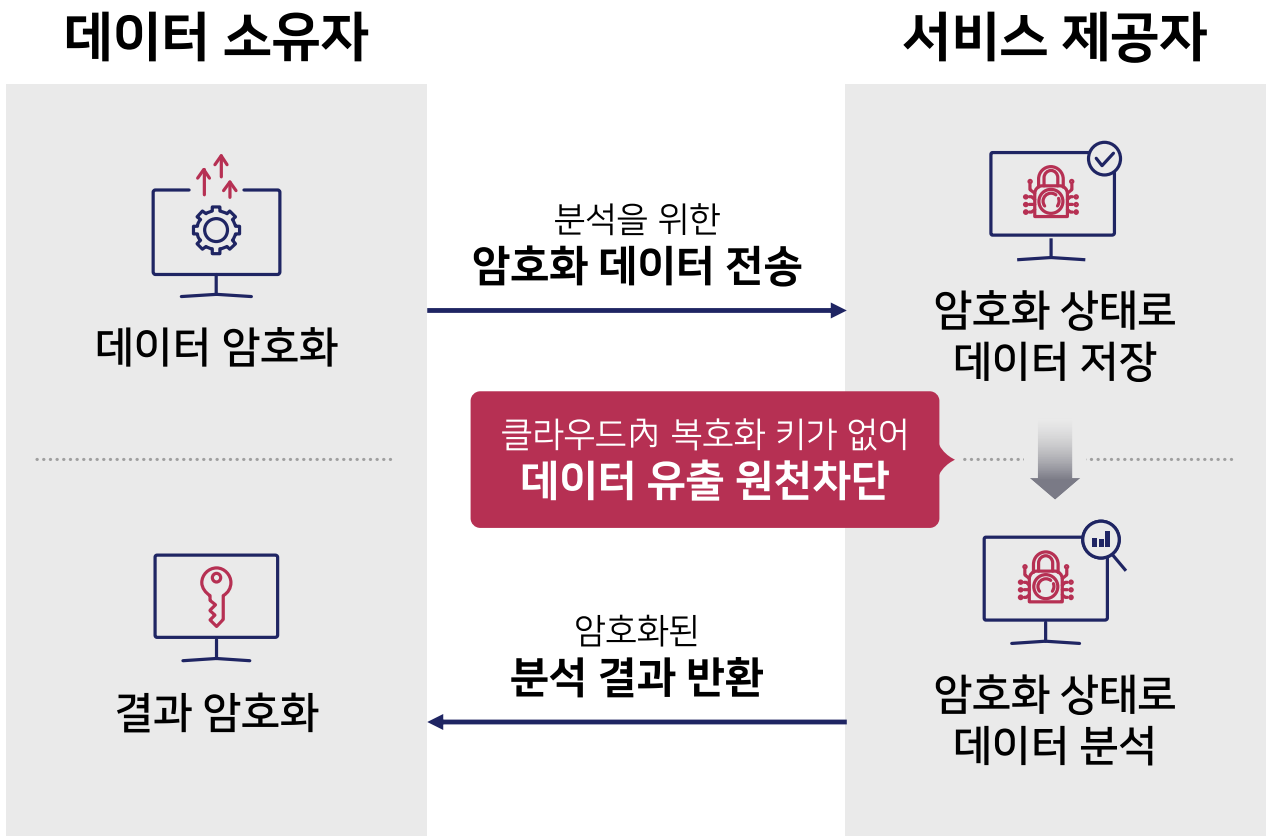
동형암호는 암호화된 상태에서 데이터 연산이 가능

주요특징

- 덧셈, 곱셈 연산 보존으로 **암호화 키를 가지지 않은 주체**도 데이터 분석 지원 가능
- 개인정보 유출이 불가함을 **수학적으로 증명 가능**

선진사

Microsoft IBM SAMSUNG SAMSUNG SDS ENVEIL ENCRYPTED VEIL Duality galois inpher

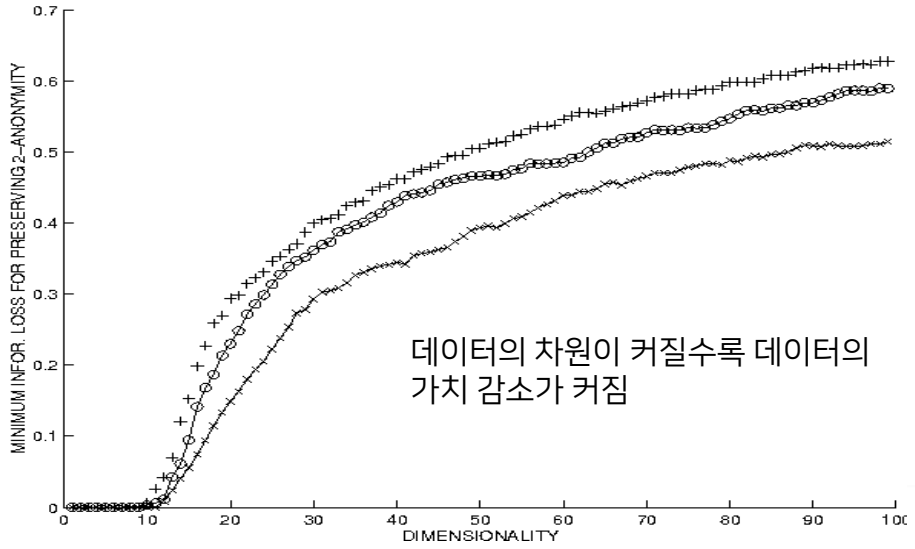


기존 기술의 한계점 – K 익명성

K-익명성 기반 비식별화 기술은 데이터손실이 크고, 프라이버시 보호에 대한 한계점 존재

01

K-익명성 기반 비식별화 時,
데이터의 차원이 커질수록 **데이터 가치 감소량 증가**



- 출처 : C. Aggarwal, On k-Anonymity and the Curse of Dimensionality

02

K-익명성을 만족하더라도 **개인정보 노출**
가능성 有 및 다른 정보와의 결합을 통한 **재식별 가능**

K-익명성 적용데이터

연령	성별	우편번호	신용등급
60대	남	180**	8
60대	남	180**	8
60대	남	180**	8
60대	남	180**	8
60대	여	180**	1
60대	여	180**	3
60대	여	180**	5
60대	여	180**	7

○○○지역의 모든 60대 남자의
신용등급은 8등급

추론 예

○○○지역에 사는 남자인
홍길동의 신용등급은 8등급

동질 집합 내에서
다양성이 부족하여
특정 개인의 정보 추론 가능

❷ 재현데이터 (Synthetic Data)

재현데이터 기술은 원본 데이터와 유사한 통계적 성질을 가지는 가짜 데이터를 생성하는 기술

- 출처: https://thispersondoesnotexist.com/

원본데이터

age	educatio	marital-	sex	hours.per.wee	income
49	HS-grad	Divorced	Female	35	<=50K
20	HS-grad	Married-civ-spouse	Female	40	<=50K
59	5th-6th	Married-civ-spouse	Male	40	<=50K

재현데이터

age	educatio	marital-status	sex	hours.per.wee	income
46	HS-grad	Divorced	Male	40	<=50K
36	Some-college	Never-married	Male	57	>50K
48	Bachelors	Never-married	Female	61	<=50K

❷ 재현데이터 (Synthetic Data)

재현데이터 기술은 원본 데이터와 유사한 통계적 성질을 가지는 가짜 데이터를 생성하는 기술

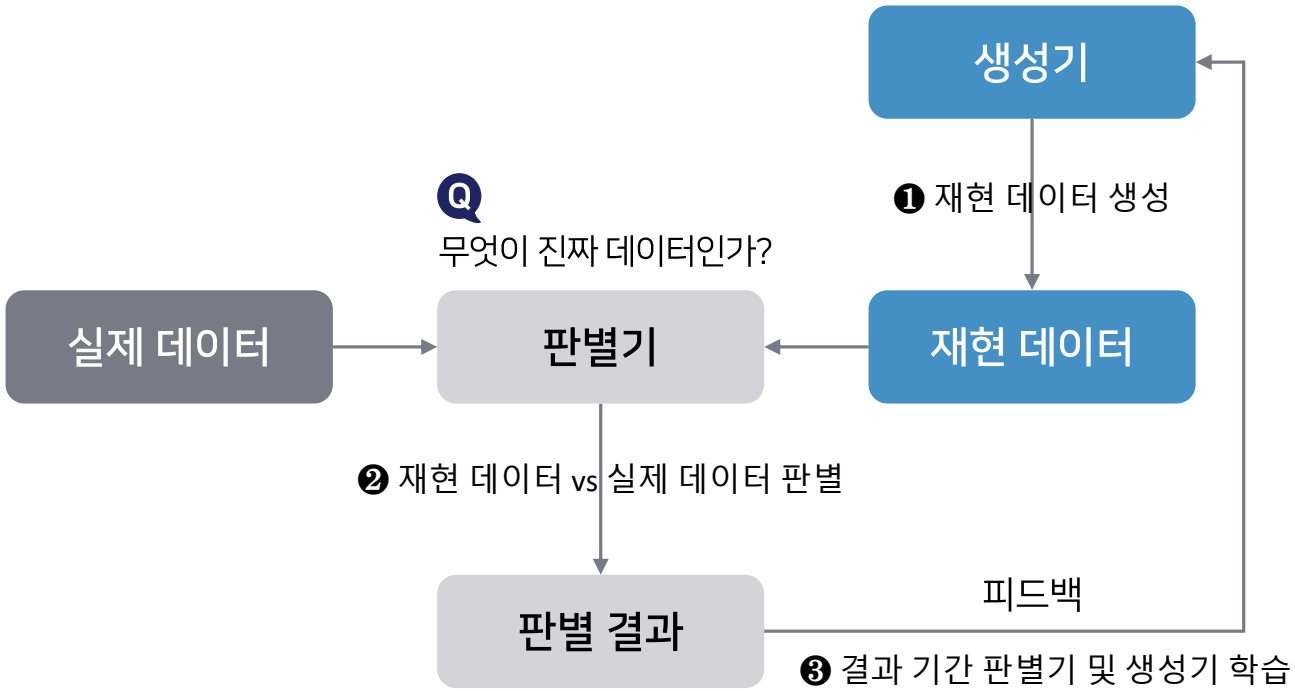
주요특징

- AI 기술 적용을 통해 원본 데이터와 유사한 통계적 성질을 가지는 재현 데이터 생성
- 원본 데이터 대신 재현 데이터를 이용한 데이터 분석 및 활용으로 프라이버시 유출 가능성 낮춤
- 차등정보보호 (Differential Privacy) 기술 적용을 통한 개인정보 보호 기능 추가 가능

선진사

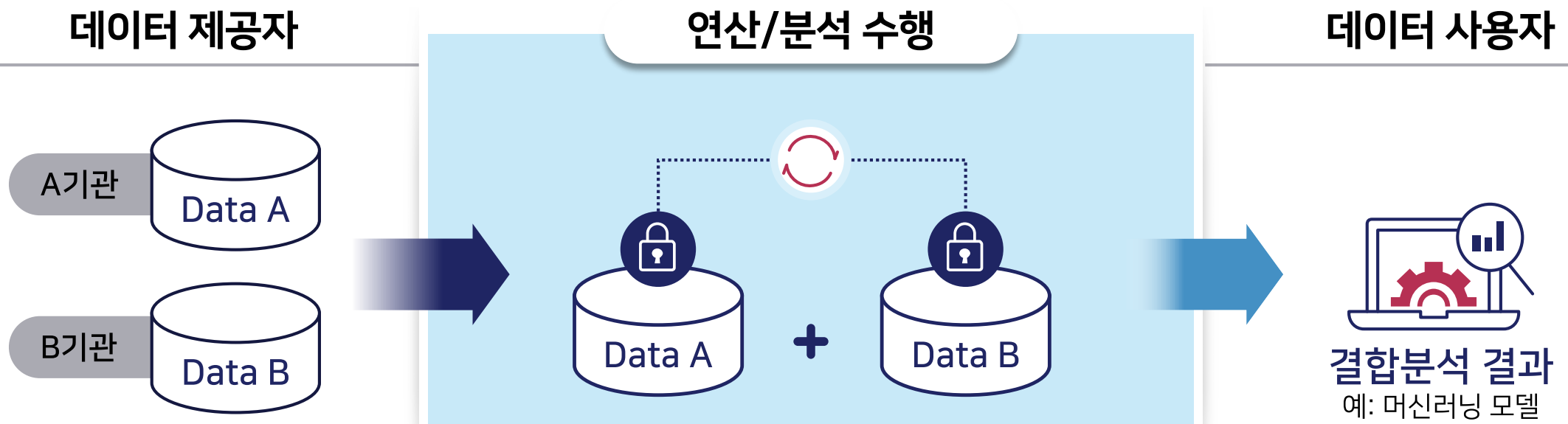
Google McAfee unity mostly AI DATAGEN

기계학습모형 (GAN) 기반 기술 적용



PET (Privacy Enhancing Tech) 소개

개인정보 유출 위험을 최소화하면서도 데이터의 가치를 최대한으로 유지하는 새로운 기술



Input Privacy

동형암호	다자간계산	기밀컴퓨팅	연합학습	재현데이터	…
Homomorphic Encryption	Multi-Party Computation	Confidential Computing	Federated Learning	Synthetic Data	

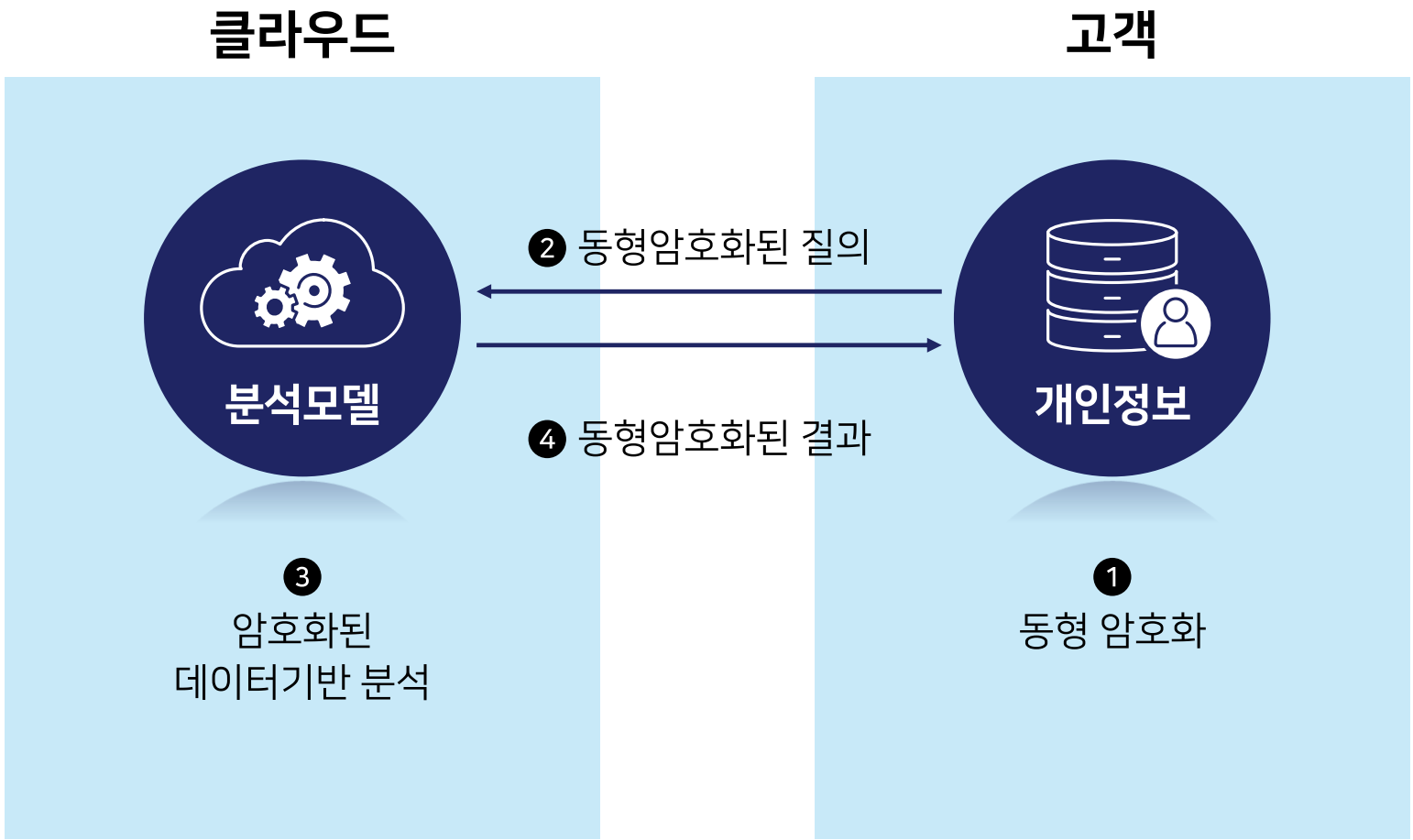
Output Privacy

차등정보보호
Differential Privacy

서비스 유형 #1 : ML/AI 기반 E2E[1] 암호화 예측 서비스

사용자 데이터와 ML 모델을 모두 보호하는 동형암호 기반 예측 Application 제공

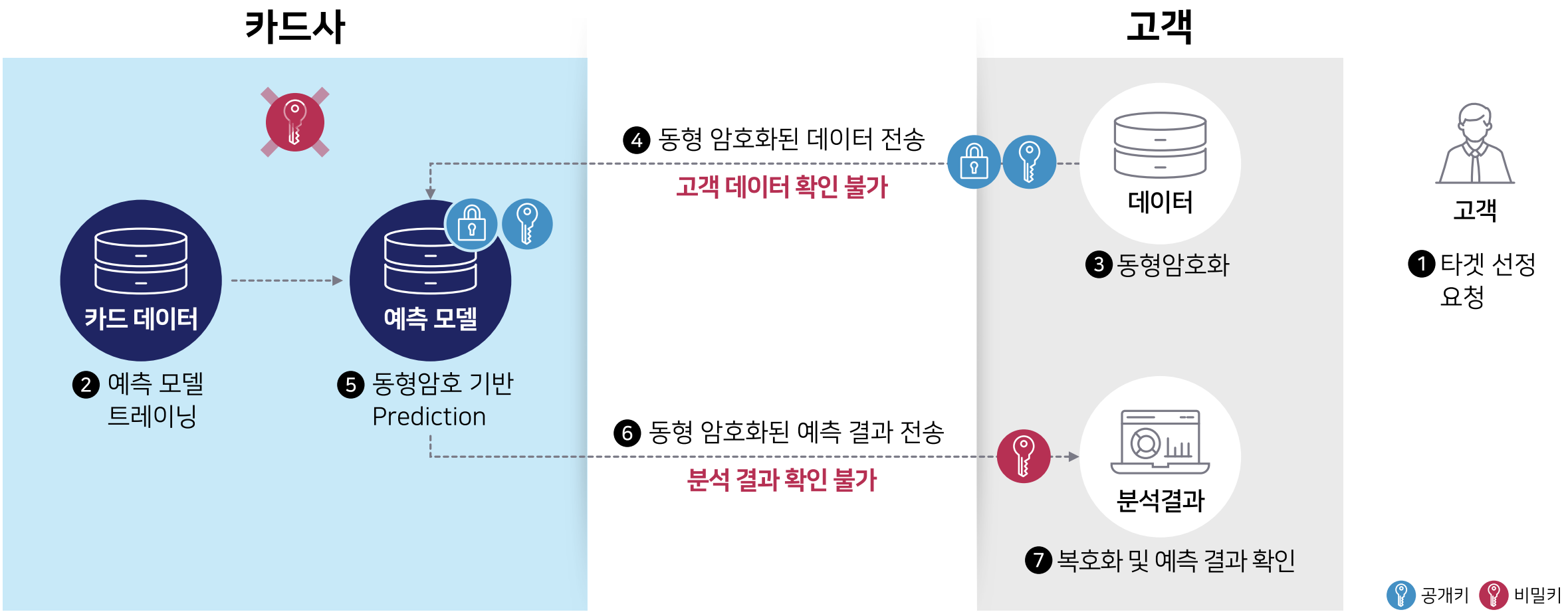
Pain Point (ML Inference)



[1] E2E : End to End

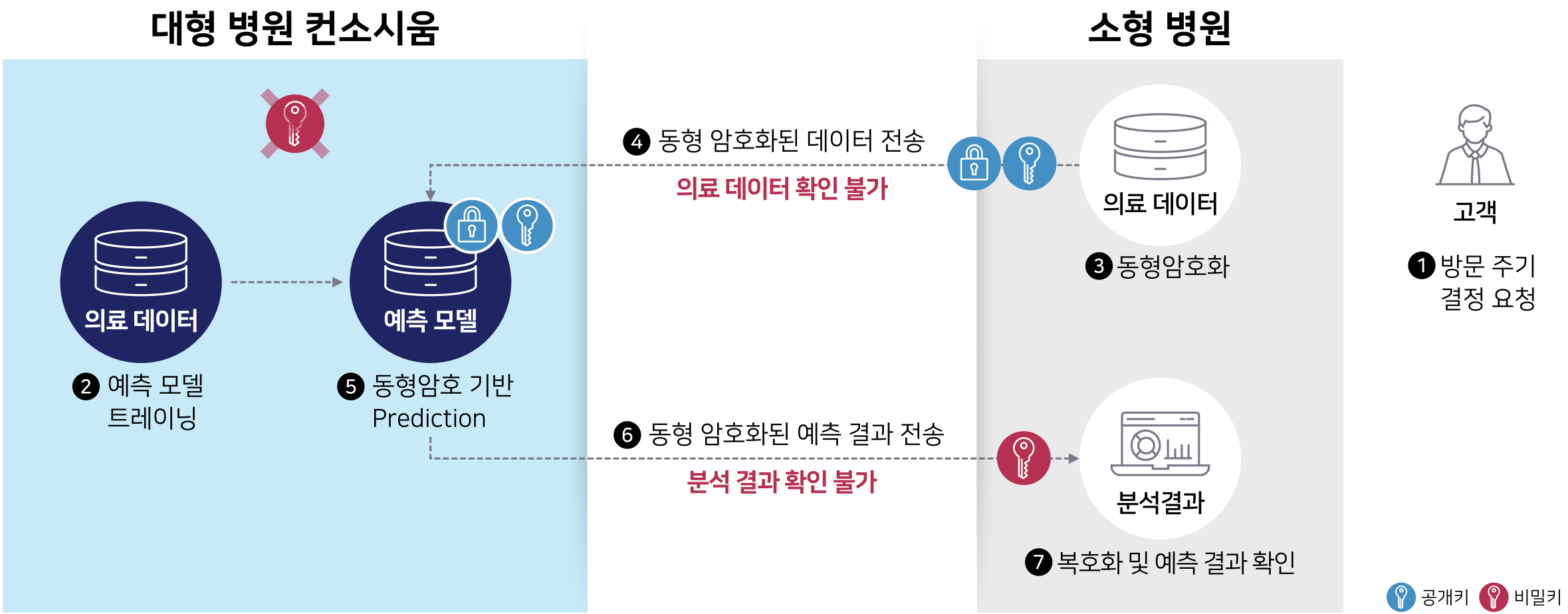
적용 사례 #1-1 : 동형암호 기반 프리미엄 고객 예측

사용자 데이터와 ML 모델을 모두 보호하는 동형암호 기반 예측 Application 제공



적용 사례 #1-2 : 동형암호 기반 유방암 재발 시기 예측

사용자 데이터와 ML 모델을 모두 보호하는 동형암호 기반 예측 Application 제공



서비스 유형 #2 : 데이터 베이스 Query 보호

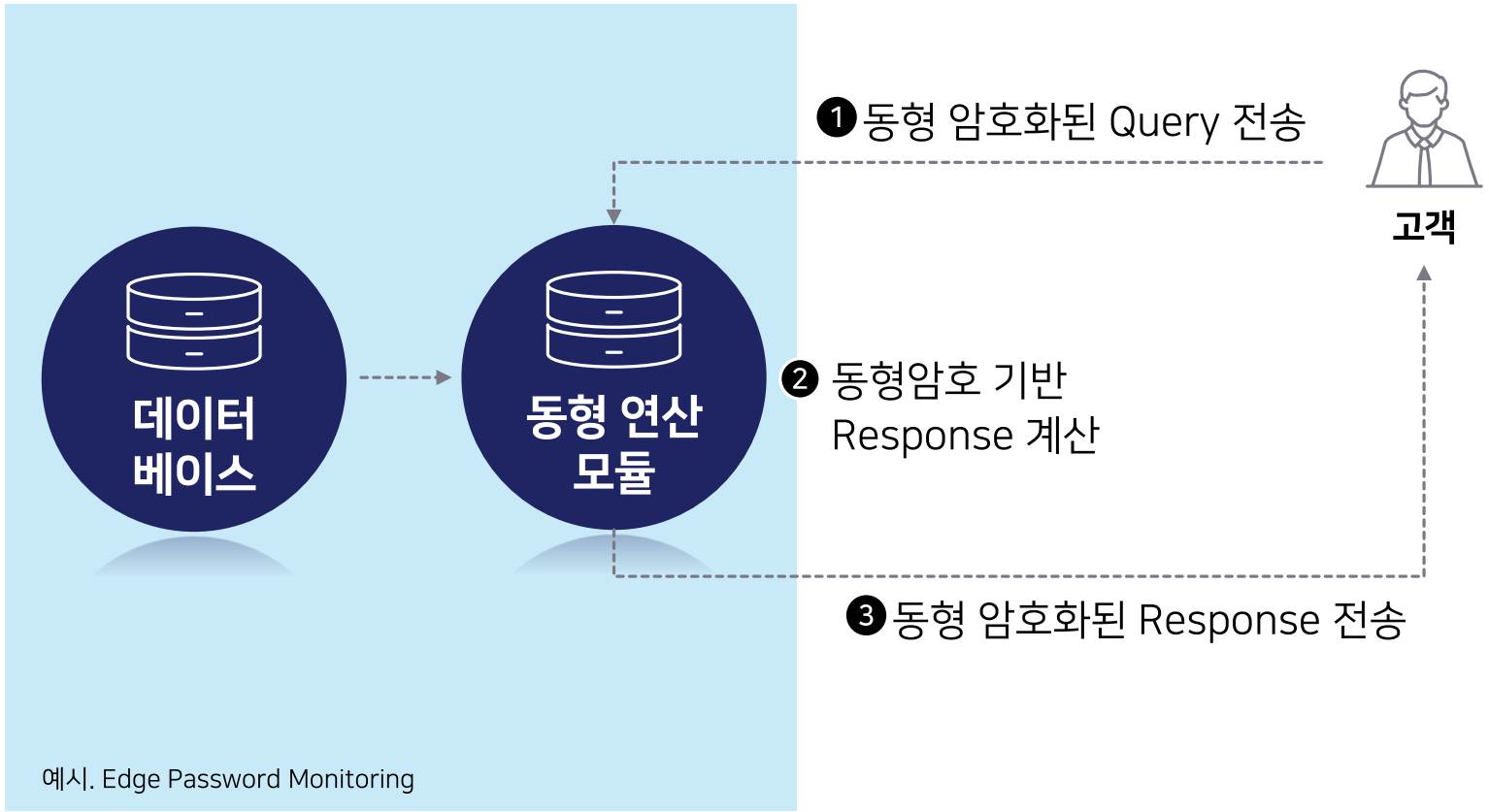
동형암호를 이용한 안전한 DB Query 보호

Pain Point (ML Inference)

사용자

Query 및 Response 정보를 통한
개인정보 유출 우려

Cloud 서비스 제공자



예시. Edge Password Monitoring

서비스 유형 #3 : 안전한 데이터 매칭을 통한 광고 효과 검증

동형암호를 이용한 안전한 데이터 매칭 및 광고 효과 검증

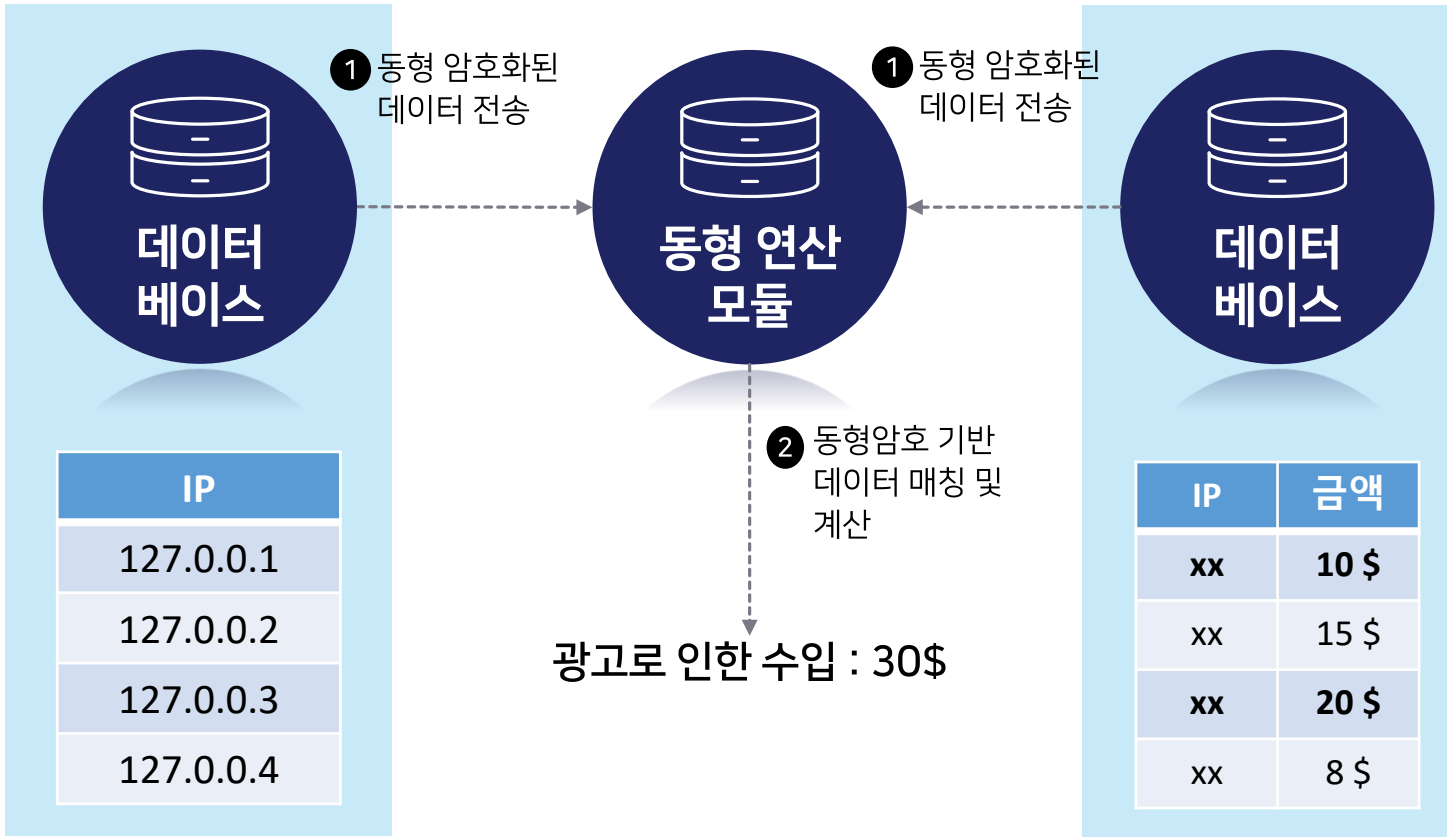
Pain Point (Matching)

사용자

고객 리스트 등 개인정보 유출 에 대한 우려

Web 서비스 제공자

온라인 쇼핑몰 운영자

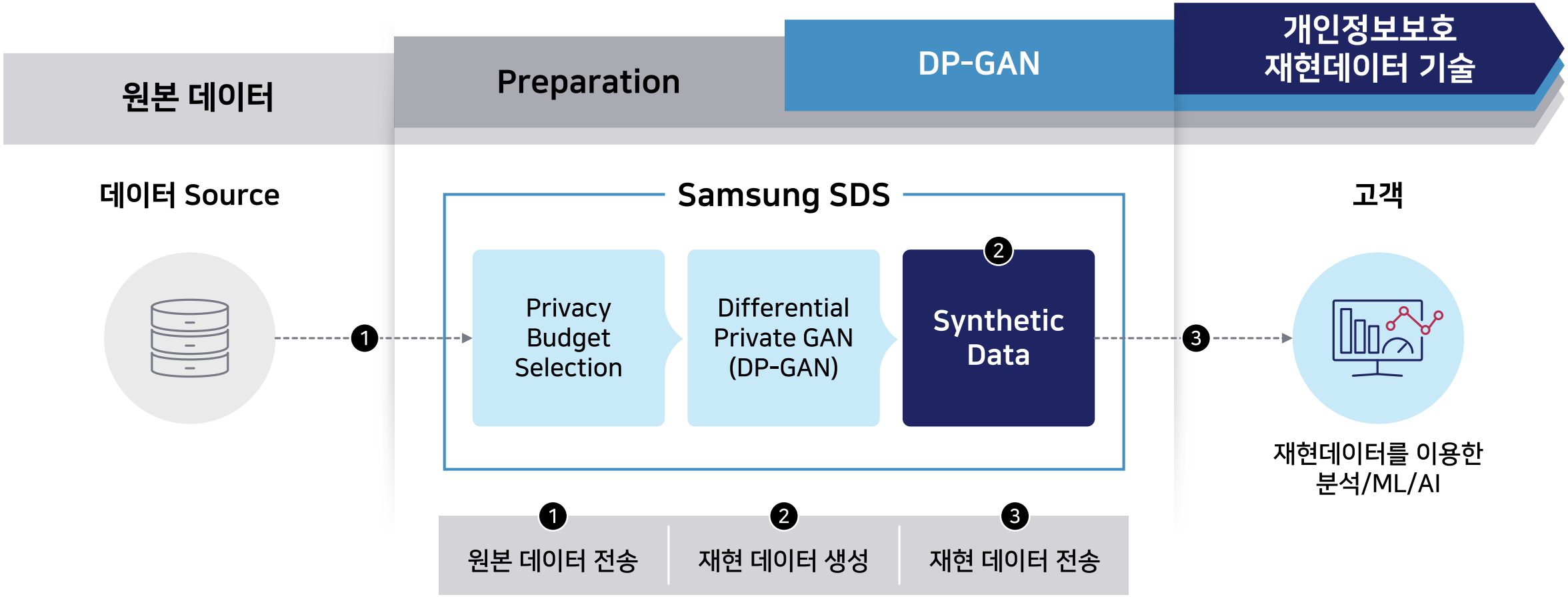


IP
127.0.0.1
127.0.0.2
127.0.0.3
127.0.0.4

IP	금액
xx	**10 $**
xx	15 $
xx	**20 $**
xx	8 $

서비스 유형 #4 : 안전한 데이터 공유 플랫폼 구축

딥러닝 기반 재현데이터 기술과 차등정보보호 기술을 활용하여 프라이버시 제어 가능한 재현 데이터 생성 및 공유



차별화 포인트

삼성 SDS는 대외 협업을 통해 세계 최고 수준의 프라이버시 강화 기술 확보 및 고도화 진행 중

대외 및 산학 협업

- 국내외 대학과의 산학 협력을 통한 차등정보보호 및 재현데이터 생성 기술 확보

독자기술 고도화 및 대외 검증

- 동형암호 데이터 분석 기술 개발 및 대외 검증
  - Cutting edge ML Training(AAAI '19) 과 동일한 정확도, 10배 이상 분석속도 제공
  - 유전 데이터 분석 국제 경진 대회 (iDASH) 우승
- 동형암호 분산 키 관리 기술 개발 및 대외 검증
  - IEEE Access 게재

국제 정보분석 보안경진대회 1위

IDASH PRIVACY & SECURITY WORKSHOP 2020

secure genome analysis competition

삼성SDS가 동형암호기술로 국제 유전체(게놈) 정보분석 보안경진대회 'iDASH 2020'에서 1위를 차지했다.

동형암호는 개인정보 등 민감한 정보를 안전하게 보호하기 위해 데이터를 암호화된 상태에서 분석/처리하는 기술이다.

'iDASH'는 2014년 미국 국립보건원(NIH)의 후원으로 시작된 전세계 유일의 유전체 정보분석 보안경진대회로, 매년 글로벌 IT기업과 대학, 연구기관들이 참가해 동형암호, 블록체인 등 보안 신기술 역량을 겨루고 있다.

삼성SDS는 올해 '동형암호 기반 암종(癌種) 분석' 부문에 출전했다. 주최측에서 제공하는 900여 명의 암호화 유전체 변이 데이터를 동형암호기술로 분석하고 암 종류를 예측하는 것이 과제이다.

전세계 36개 팀이 참여한 이번 대회에서 삼성SDS는 최고 수준의 분석 속도와 정확도를 기록하며 서울대학교를 비롯한 2개 보안전문기업과 함께 공동 1위에 선정됐다.

Thank you

SAMSUNG SDS