

**SAMSUNG SDS**

Realize your vision

# Techtonic 2019

Partner



Foresee

Disrupt

2019.11.14 • SAMSUNG SDS Tower B1F  
{ Magellan Hall / Pascal Hall }

## Track 3 | Security

# 정보 손실/유출 없이 고객 데이터를 분석해보자!

문덕재 프로 (보안알고리즘Lab) / 삼성SDS  
김동우 박사과정 (Crypto Lab) / 서울대학교

# AGENDA

1. 프라이버시관련 동향
2. 동형암호기술
3. 동형암호 적용사례
4. 클라우드 적용시 고려사항

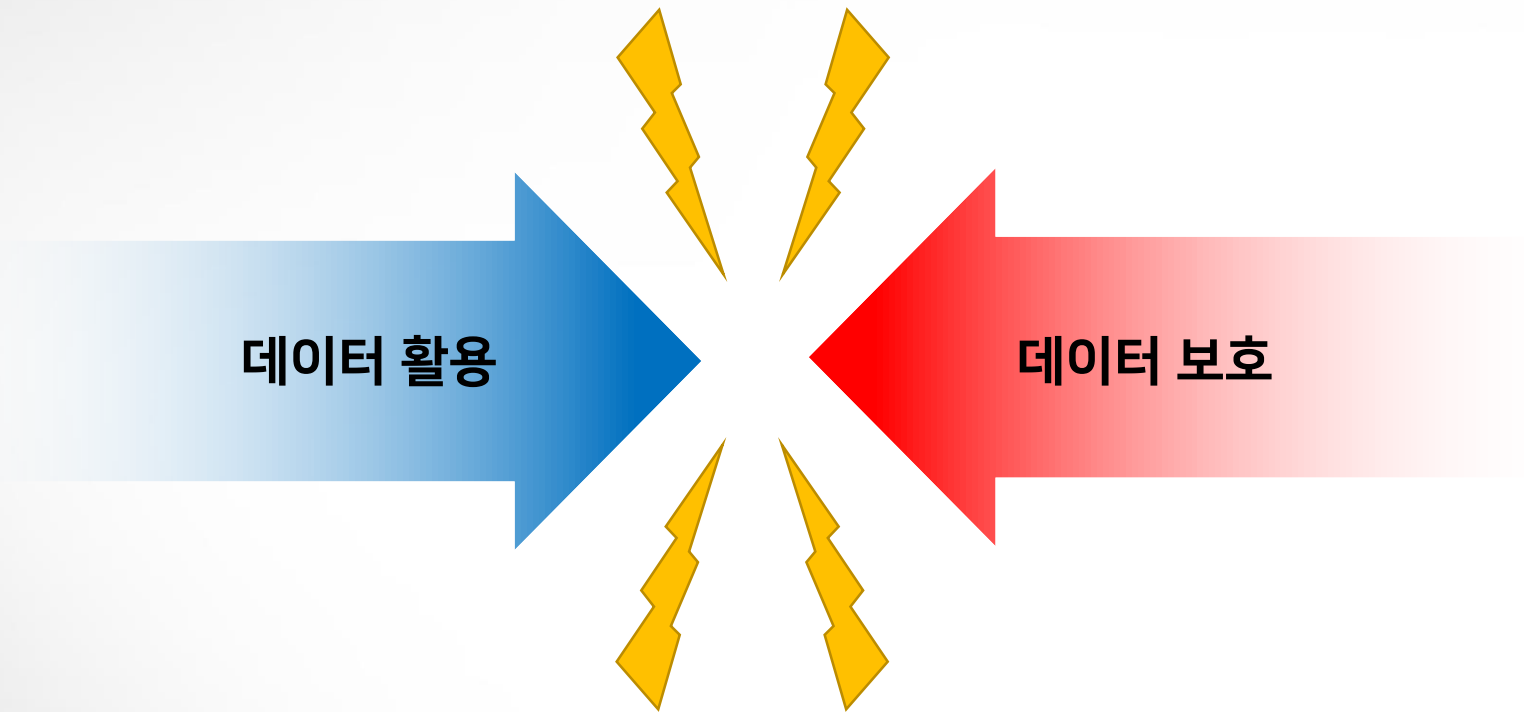
1

---

# 프라이버시관련 동향

---

# 들어가며



데이터가 중심이 되는 세상 도래에 따라 글로벌 GDPR(EU), CCPA(US), PIPEDA(CA), 국내 개인정보보호법 등의 법/규제들이 **데이터 보호와 활용간 전쟁을 중재하기 시작함**

# 국제표준 비식별화 기술

국제표준 ISO/IEC 20889 (IS\*, 2018) 에서 비식별화 기술 표준화 완료

## 주요 비식별화 기술

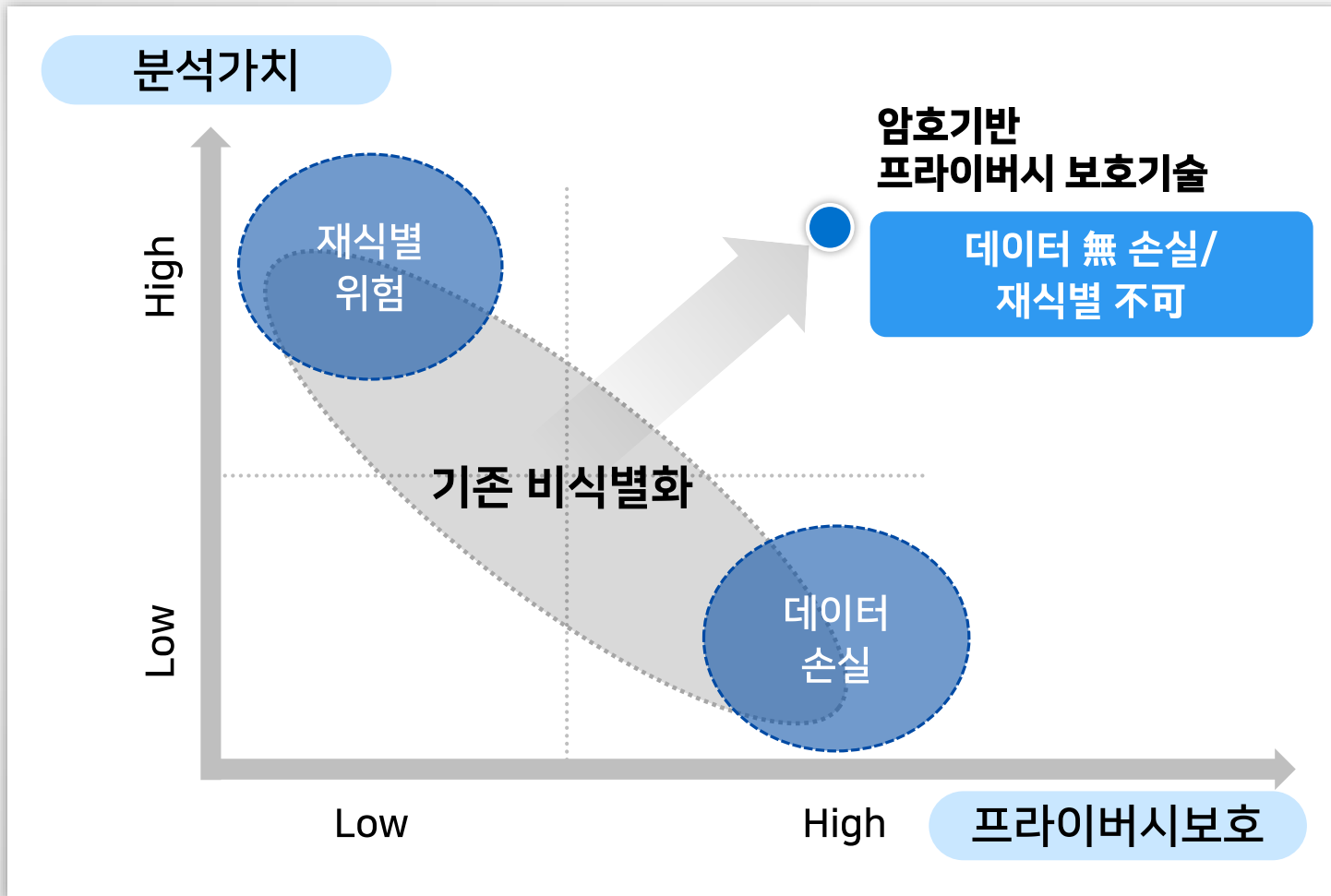
- 1 Pseudonymization - Hashing
- 2 Generalization - Rounding
- 3 Randomization - Noise addition
- 4 **Cryptographic tools**
  - Homomorphic Encryption
  - Homomorphic Secret Sharing

#	성명	연령	주소	방문일
1	홍길동	21	서울 서초구 성촌길 56	'17.12.21
2	김잔디	35	전라남도 여수시 이상 5로	'18.5.23
3	전우치	54	서울 도봉구 방학로 4길	'18.6.2
4	장보고	29	서울 강동구 조정대로 43번길	'18.6.2
5	장희진	36	전라남도 여수시 갈월 5길	'18.10.31

#	성명	연령	주소	방문일
1	0111...1010	20대	서울 서초구	'17.12.23
2	1001...1110	30대	전라남도 여수시	'18.5.25
3	0001...0010	50대	서울 도봉구	'18.6.4
4	0011...1001	20대	서울 강동구	'18.6.3
5	0011...0010	30대	전라남도	'18.10.29

# 프라이버시 보호기술의 변화

기존 프라이버시 보호기술의 한계로 암호기반 프라이버시 보호기술 연구 및 사업적용 요구 증가



“ Microsoft, Intel, SAP, Google, ANT Financial 등이 암호기반 프라이버시 보호기술 도입 진행 중 ”

암호기반 프라이버시 보호기술

1. 동형암호
2. 다자간계산
3. 차등정보보호

2

---

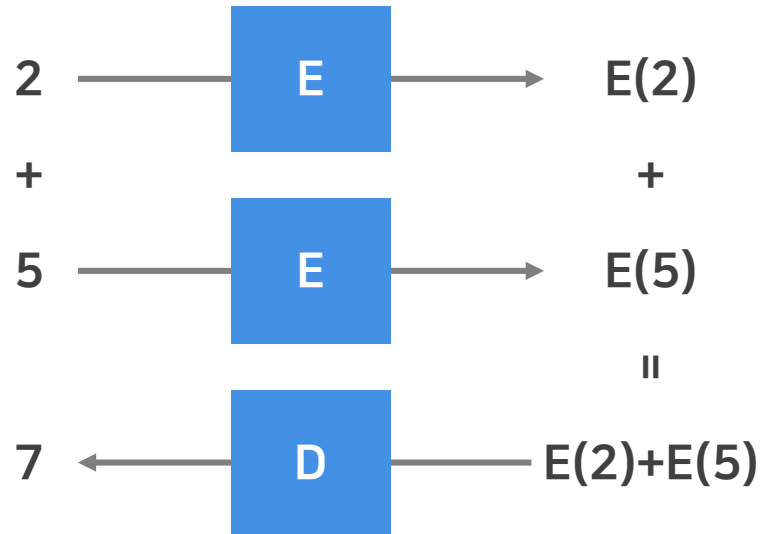
# 동형암호기술

---

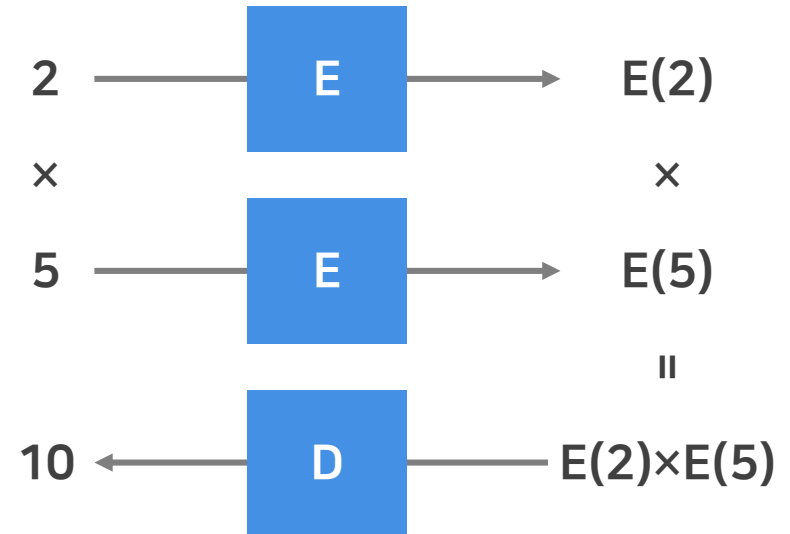


# 동형암호란?

## 덧셈 연산 보존



## 곱셈 연산 보존



동형암호는 덧셈/곱셈 연산을 보존하여,  
암호화된 데이터의 연산 결과를 복호화하면 원본 데이터의 연산 결과와 동일

\* 동형(Homomorphic): 동일한 유형의 두 대수 구조 사이의 연산을 보존하는 사상을 의미하는 동형성(Homomorphism) 에서 유래됨

# 일반 암호화 vs. 동형 암호화

## 일반 암호화된 데이터 분석



일반암호화 데이터

## 동형암호화된 데이터 분석



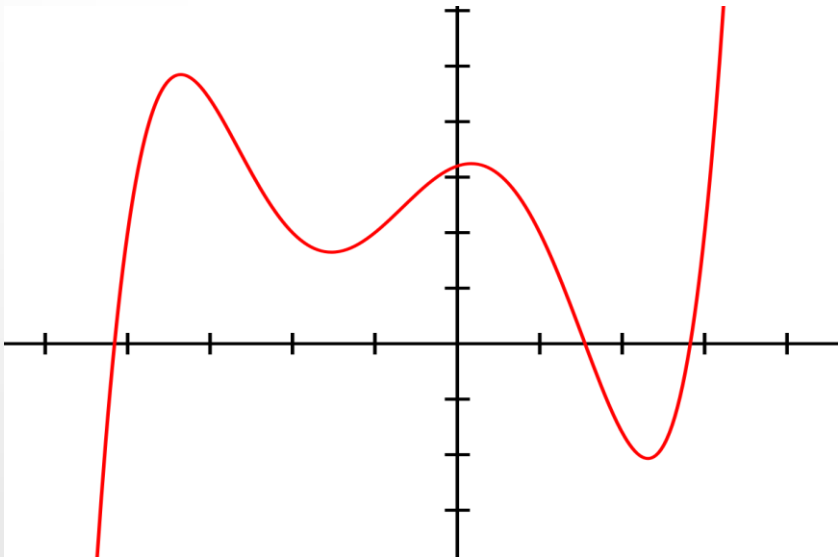
동형암호화 데이터

# 동형암호기반 암호화된 데이터분석

데이터의 손실/유출 없이 암호화된 상태에서 데이터분석 (머신러닝/딥러닝) 이 가능  
(머신러닝과 딥러닝 분석함수들은 **모든 산술연산으로 표현 가능**)

## 親 동형암호 연산

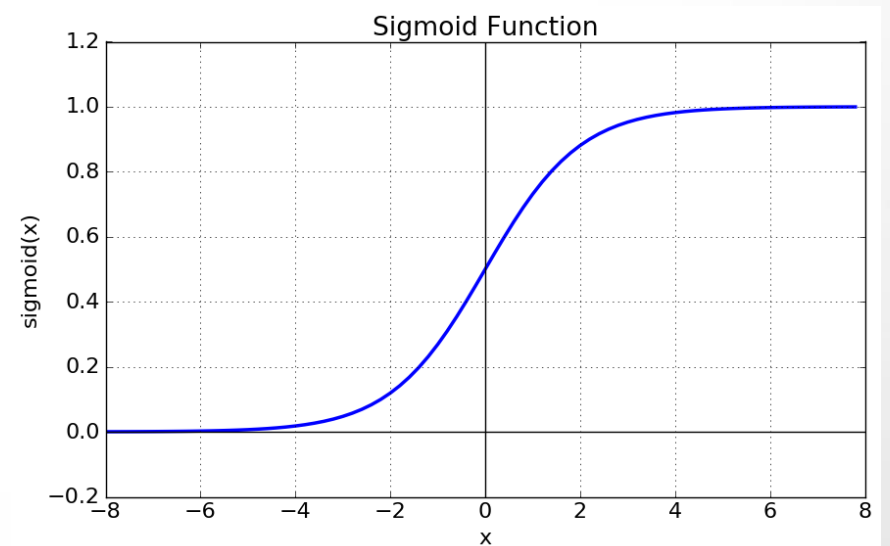
- ▶ 산술연산들 중 덧셈/곱셈 연산의 조합으로 표현 가능한 **親 동형 암호연산**(예, 다항식 연산) 의 경우 **동일한 분석 정확도를 제공**



Source: [https://ko.wikipedia.org/wiki/오차\\_방정식](https://ko.wikipedia.org/wiki/오차_방정식)

## 非親 동형암호 연산

- ▶ 덧셈/곱셈 연산의 조합으로 표현 불가능한 연산들 (예, 지수/로그, 비교함수 등) 의 경우 **근사식**을 통해 표현가능하나 **근사정도에 따라 정확도에 차이가 존재**



Source: [https://en.wikipedia.org/wiki/Sigmoid\\_function](https://en.wikipedia.org/wiki/Sigmoid_function)

# 삼성SDS 동형암호기술 로드맵

세계 최고의 동형암호기술기반 데이터분석분야 신규 사업기회 마련을 위한 분석기술 개발

## Privacy Enhancing Technology

### Service

금융

의료

마케팅

제조

### Application Software

Statistic Analysis

Machine Learning

Deep Learning

### Primitive Software

HeaAn Lib.

SEAL Lib. (Open SW)

### Computing Resource

CPU

고속화

### Computing Environment

On-Prem.

Cloud

## 삼성 SDS 는 세계 최고의 동형암호기술 (HeaAn) 확보를 위해 서울대학교 암호랩과 협업

### ▶ 학계 검증 완료

- Asiacrypt 2017, Eurocrypt 2018, SAC 2018 등 Top-tier 암호학회 발표를 통해 학계 검증 완료

### ▶ 암호화된 분석지원 국제대회 검증 완료

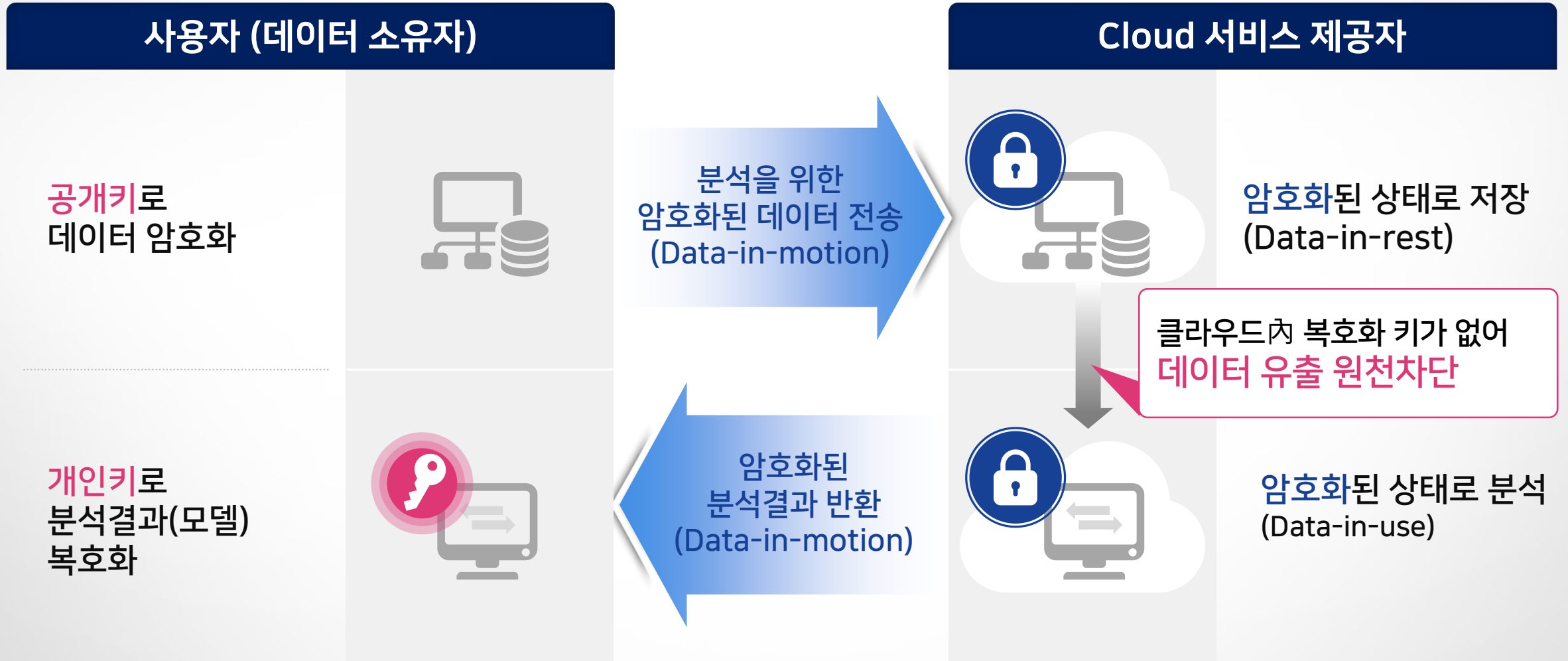
- 2017 iDash 우승
- 2018 iDash 모든 예선 통과자들이 HeaAn 사용

### ▶ 분석속도 및 정확도 비교 (2017 iDash 기준)

순위	참여사	분석속도	오차
1	서울대 (당사 산학)	10분	1%미만
2	M社 (미국)	6시간(36배)	5%미만
3	C研 (프랑스)	36시간(216배)	1%미만

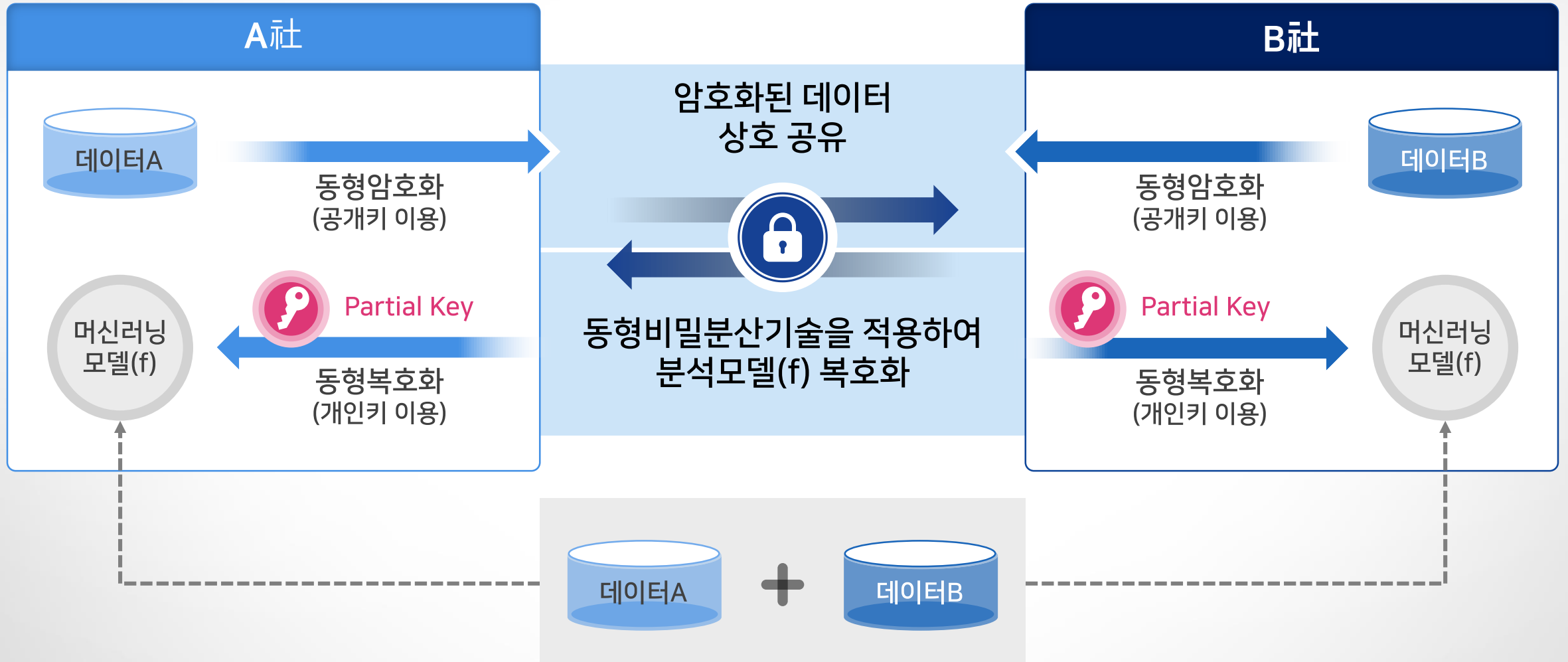
# Use Case(1/3) | Learning(사용자 데이터 분석지원)

사용자가 데이터 유출 걱정없이 퍼블릭 클라우드 등의 외부 분석 서비스를 이용할 수 있는 환경 제공



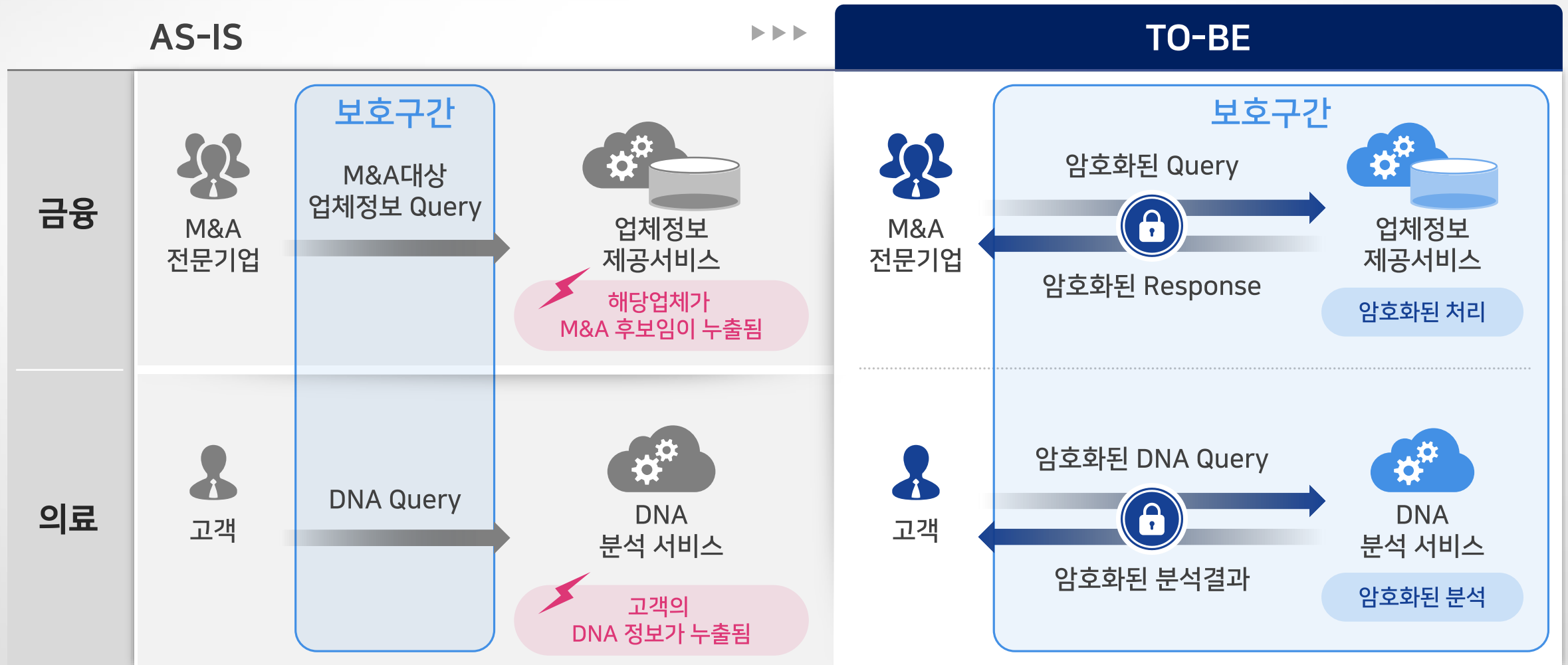
# Use Case(2/3) | Learning(분산 데이터 결합분석지원)

데이터 유출 걱정없이 2개 이상의 데이터 주체로부터 데이터 결합이 가능하여 더 정밀한 분석모델 도출 가능



# Use Case(3/3) | Prediction(고객 질의/응답내용 보호)

질의하는 데이터의 유출없이 원하는 분석 서비스 이용 가능



3

---

# 동형암호 적용사례

---



# 기술PoC(1/2) | 삼성카드사 카드거래 데이터

동형암호화된 카드거래 데이터 기반 프리미엄카드 신청자 분석 모델 개발 및 예측 PoC

## PoC 개요

PoC 명

동형암호 기반  
프리미엄카드 신청자  
분석 모델 개발 및 예측

PoC 목표

원본 데이터와 암호화된 데이터間  
분석 모델 개발 및 예측에 대한  
정확도와 성능 비교

분석함수

Logistic Regression

## 분석 데이터 (삼성카드 제공)

1. 데이터 개요

- . 전체 데이터: 1,001,153건
- . Feature 수: 125종

2. 데이터 셋 구성

- . 데이터 제공형태: CSV 파일
- . 모델 개발용 데이터 셋: 700,807건 (미신청 700,013건, 신청 794건)
- . 모델 검증용 데이터 셋: 300,346건 (미신청 299,987건, 신청 359건)

# 기술PoC(2/2) | 삼성카드사 카드거래 데이터

카드 거래내역 데이터를 사용한 분석(Logistic Regression) 수행결과 확인

- 수행 결과: 원본데이터와 동형암호화된 데이터간 분석 정확도(AUROC) 일치

• Sample 개수: 300,346개

구분	SDS 분석함수		카드 분석함수
	원본 데이터	동형암호화된 데이터	Python (삼성카드 테스트)
AUROC	0.7717	0.7722	0.8357

- SDS 분석함수 계수비교 : 대체로 같음

• Sample 개수: 700,807개 / Feature 개수 : 125개

SDS 분석함수	원본 데이터	0.0159	0.0095	...	-0.0503	-0.0191	...	-0.3446
	암호화된 데이터	0.0159	0.0095	...	-0.0527	-0.0201	...	-0.3534

# 4

---

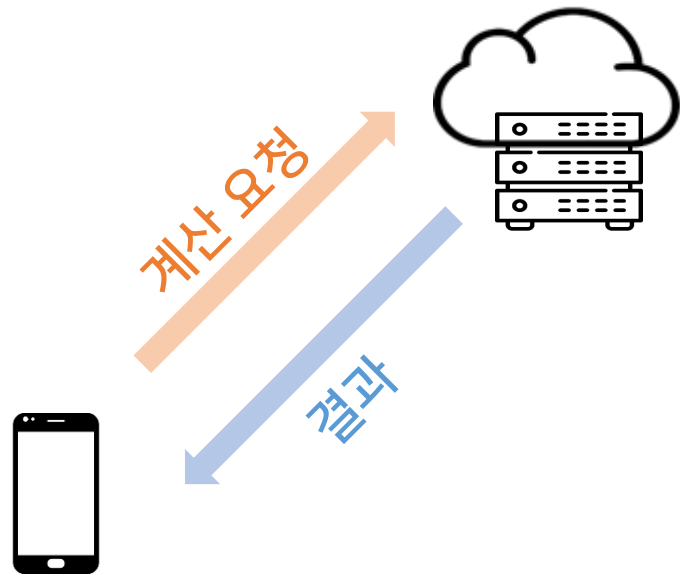
## 클라우드 적용시 고려사항

---

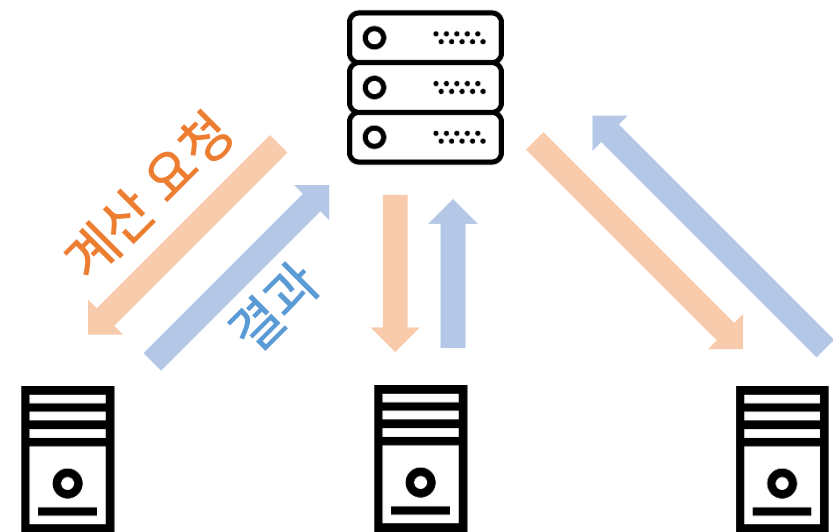
# 배경 | 계산 외주 (Outsourced Computation)

신뢰할 수 없는 기관에 중요한 계산 요청시 결과를 어떻게 효율적으로 검증할 수 있을까?

## 클라우드 컴퓨팅



## 분산 컴퓨팅



# 해결방안 | 계산 검증 기술 (Verifiable Computation)

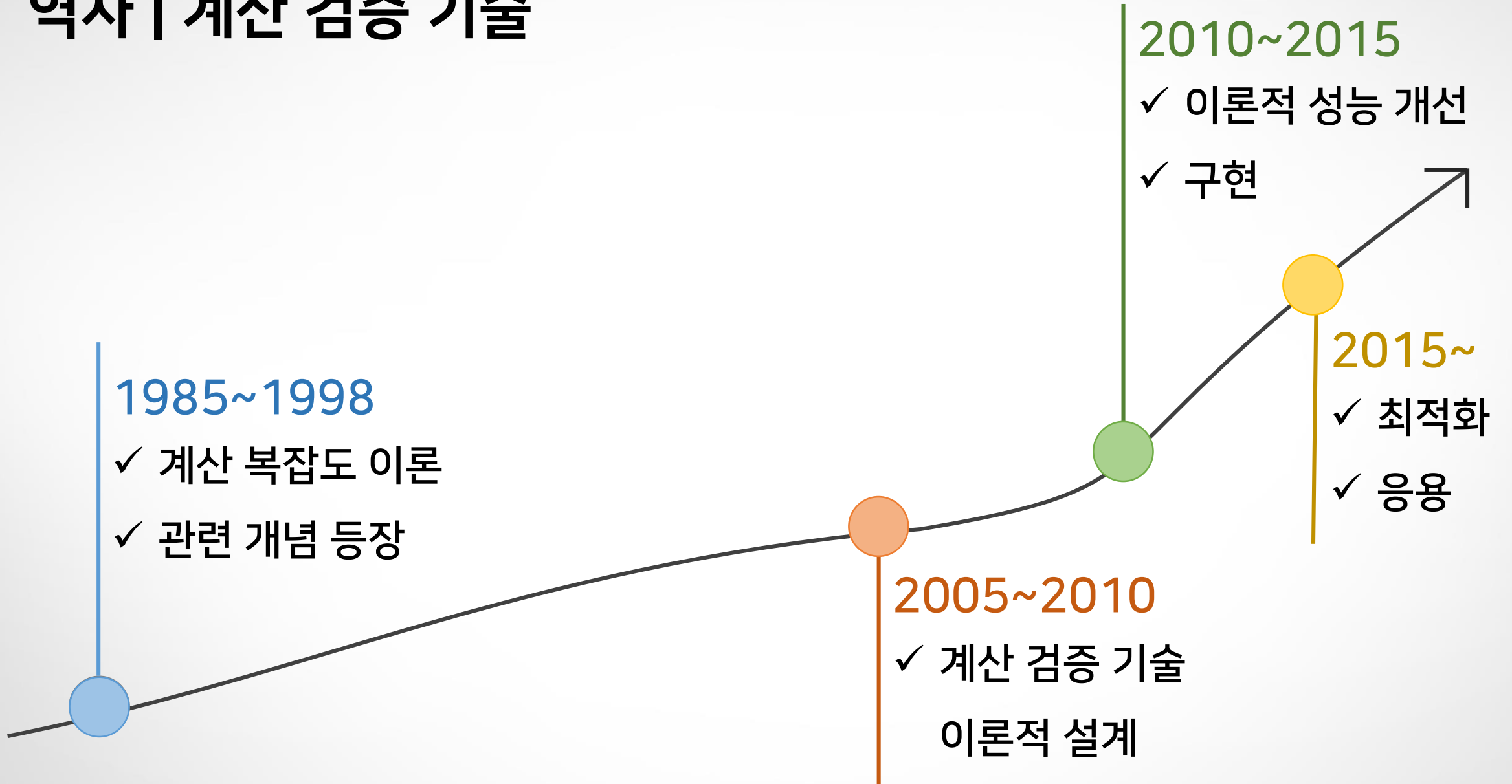


(증명 & 결과) 검증 비용  
**<< 계산 비용**



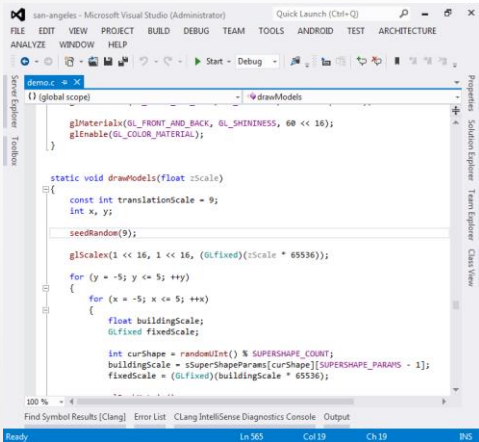
계산 비용  
**<< 증명 생성 비용**

# 역사 | 계산 검증 기술



# 프로세스(1/2) | 계산 검증 기술

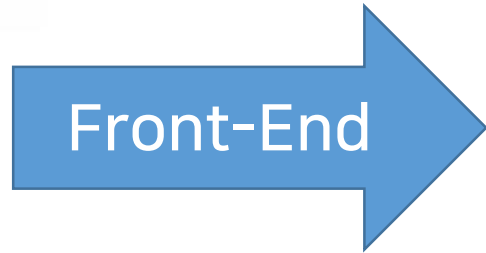
프로그램  
(C/C++, JAVA, Python...)



```
gMaterial(0L_FRONT_AND_BACK, 0L_SHININESS, 00 << 16);
gEnable(0L_COLOR_MATERIAL);

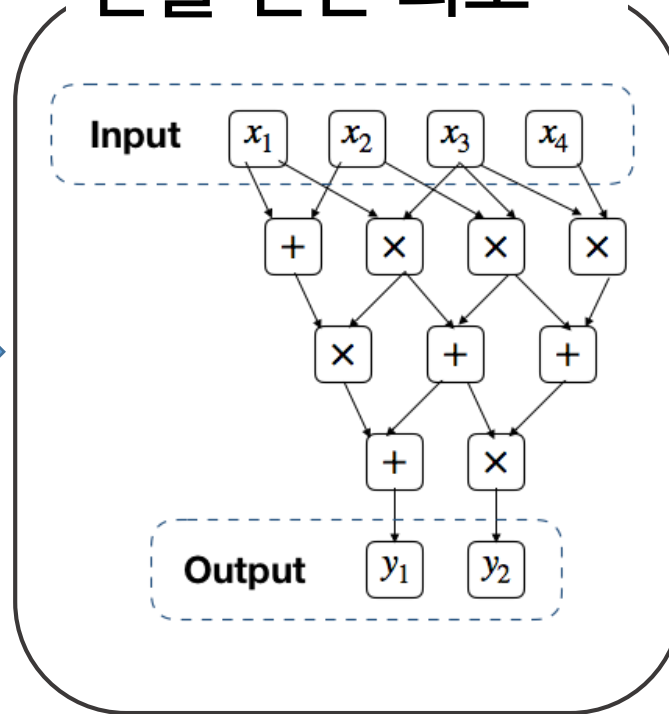
static void drawModels(float zScale)
{
    const int translationScale = 9;
    int x, y;
    seedRandom(9);
    glScalex(1 << 16, 1 << 16, (0LFixed)(zScale * 65536));
    for (y = -5; y <= 5; ++y)
    {
        for (x = -5; x <= 5; ++x)
        {
            float buildingScale;
            0LFixed fixedScale;

            int curShape = randomInt() % SUPERSHAPE_COUNT;
            buildingScale = sSuperShapeParams[curShape][SUPERSHAPE_PARAMS - 1];
            fixedScale = (0LFixed)(buildingScale * 65536);
```



연산 비용 증가  
 $1x - 10,000x$

산술 연산 회로



증명 검증

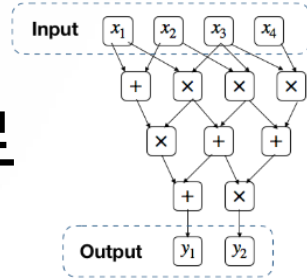


증명 생성

비용  $10x - 1,000x$

# 프로세스(2/2) | 계산 검증 기술

프로그램 → 산술 연산 회로



Key 생성 알고리즘



증명키

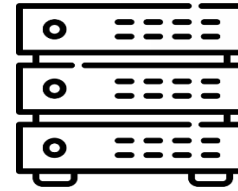


검증키



증명키

입력



서버



증명

연산결과



검증키



의뢰인

정확하다!

이상하다?



# 성능 | 계산 검증 기술

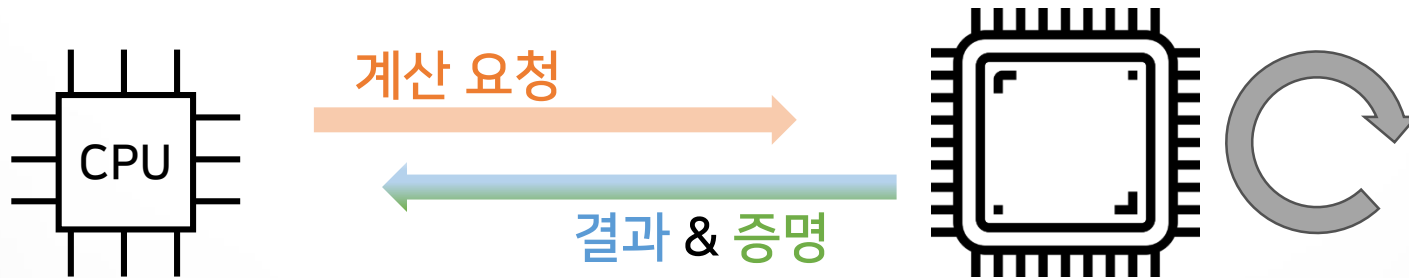
- ✓ Key생성 비용  $\propto$  (프로그램) ~ 0.1 |C| ms
- ✓ 증명 생성 비용  $\propto$  산술 연산 회로 크기
- ✓ 증명 검증 비용  $\propto$  입/출력 크기 ~ 5 + 0.001 |N| ms
- ✓ 증명 크기 일정 288 Byte

연산	증명 시간	시간 (산술 회로)	검증 시간	연산 시간 (프로그램)
행렬곱 (128 × 128)	1200 sec	510 ms	15 ms	16 ms
해쉬 (SHA-1) 계산	16 sec	19 ms	10 ms	1 us

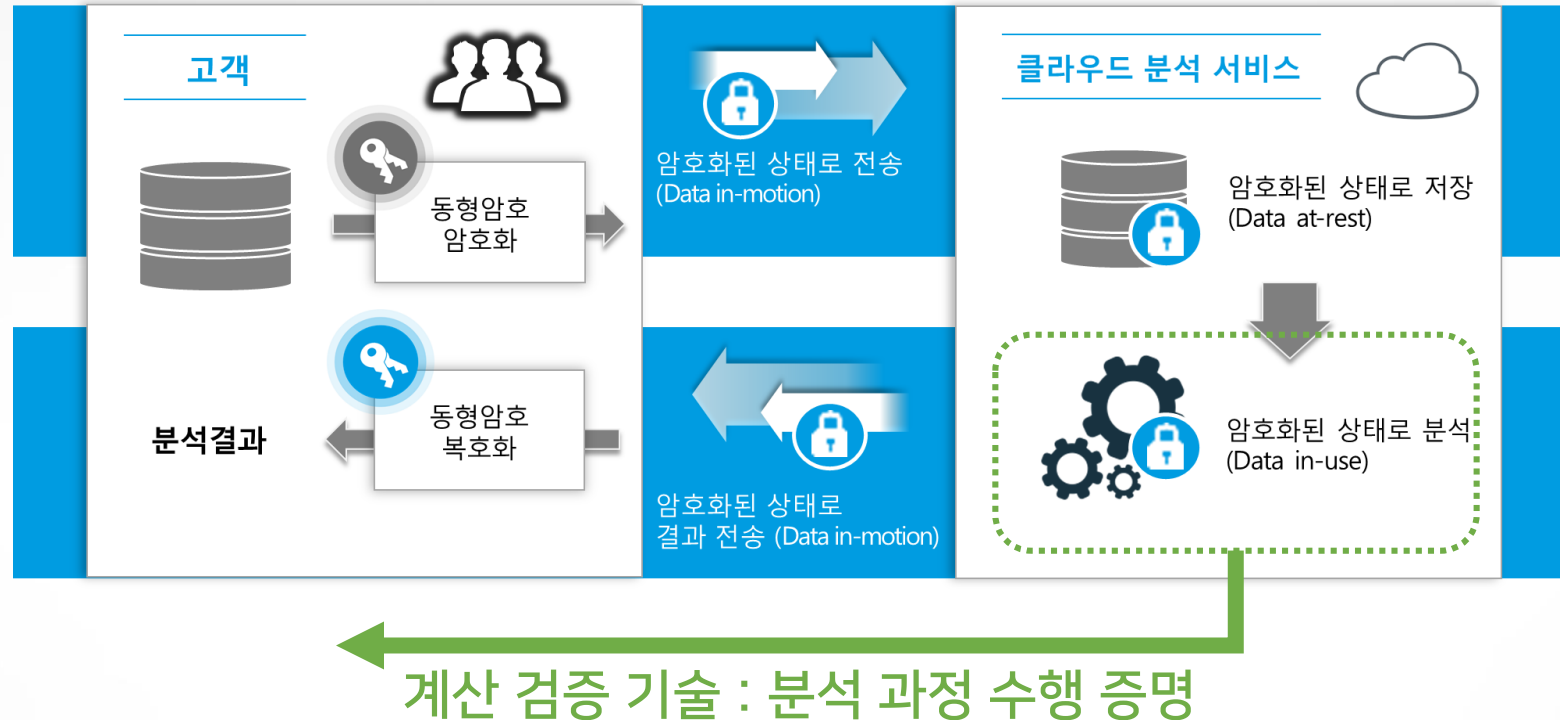
Ref) Ben-Sasson, Eli, et al. "Succinct non-interactive zero knowledge for a von Neumann architecture." *USENIX Security 14*. 2014. & updated version 2019.  
 Parno, Bryan, et al. "Pinocchio: Nearly practical verifiable computation." IEEE Symposium on Security and Privacy 2013.

# 응용 | 계산 검증 기술

- ✓ 클라우드 컴퓨팅 / 분산 컴퓨팅 : 신뢰성 보장
- ✓ 고성능 하드웨어의 수행 검증 [S&P'16, CCS'17]

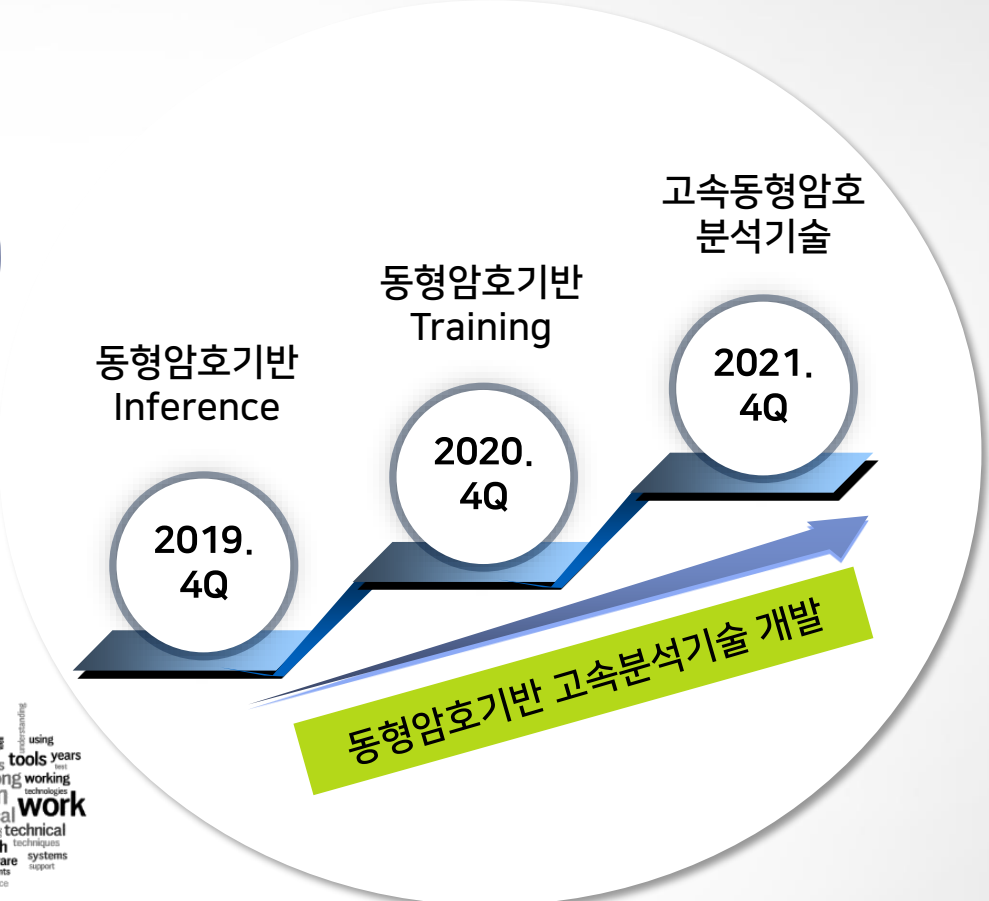


# 동형암호기술 연계 | 계산 검증 기술



- ✓ 데이터 프라이버시와 계산의 신뢰성을 동시에 보장!
- ✓ 성능의 한계 극복 필요

# 마치며



다양한 사업분야에서 요구하는 분석함수들에 신규 프라이버시 보호기술을 적용하여  
 고객의 데이터를 안전하게 분석할 수 있는 **세계최고의 프라이버시 강화 분석기술을 제공**



**Thank You**



The graphic features the text 'Q & A' in a clean, sans-serif font. The 'Q' and 'A' are white, while the ampersand is a vibrant lime green. To the right of the 'A', two orange triangles point rightward, one slightly overlapping the other, leading towards a large, dark blue circle on the far right. The background is a dark blue gradient with a pattern of lighter blue circles on the left side.

Q & A

Partner Disrupt Foresee