



1. 4차 산업 혁명의 뿌리 기술
2. 블록체인과 데이터
3. 데이터 기술 유형
4. 데이터 산업 적용을 위한 5가지 블록체인 특징
  - 데이터에 대한 강력한 믿음 '신뢰 보장(데이터 무결성)'
  - 데이터 해킹 No '악성 행위 방지'
  - 정확한 데이터 분석 '예측 분석'
  - 데이터 즉시성 강화 '실시간 데이터 접근'
  - 모두의 데이터 '데이터 공유 관리'
5. 데이터 산업 적용 가능 블록체인 유스케이스
  - 개인정보 비식별화 > 처리방안
  - 자율 데이터 운영관리 강화
6. 별첨

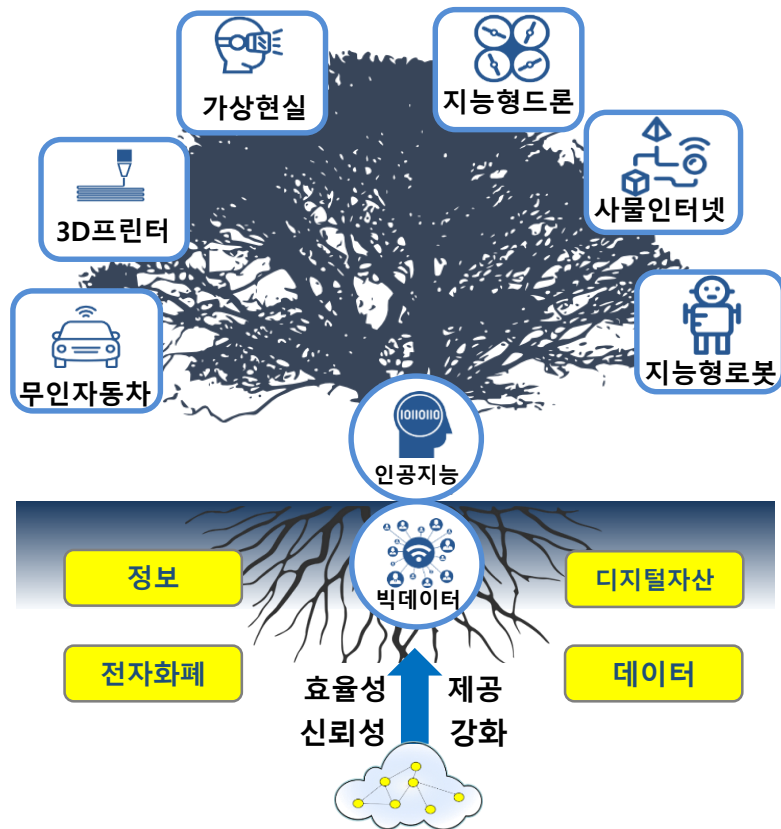
## 1. 4차 산업 혁명의 뿌리 기술

### 블록체인

빅데이터와 함께  
4차 산업혁명의 뿌리  
핵심 인프라

데이터 및 자산거래의 신뢰성을 제공하여  
트랜잭션 처리 비용의 효율성을 제고

## 4차 산업혁명 신성장 산업

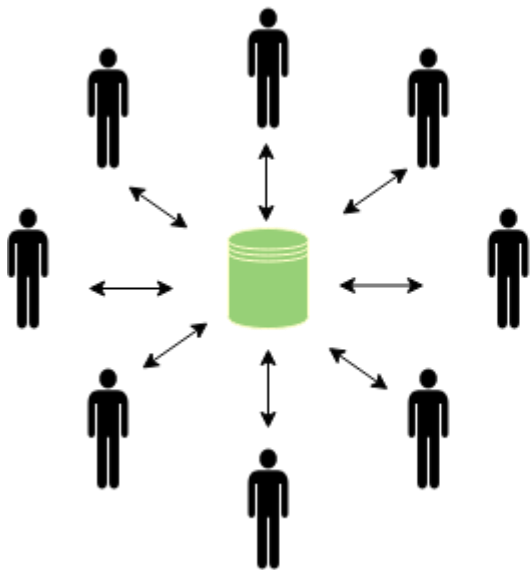


(고려대 인호교수님,  
"블록체인: 4차 산업 혁명의 혁신 인프라" 자료 인용)

블록체인 기술

## 2. 블록체인과 데이터

중앙집중형 데이터베이스  
(Centralized)



VS

분산형 데이터베이스  
(Distributed)



## 2. 블록체인과 데이터

### 블록체인의 기본 구조



### 블록체인의 거래 흐름



## 2. 블록체인과 데이터

빅데이터 : 블록체인 = Quantity : **Quality**

### 빅데이터와 블록체인의 융합

#### Security

다수가 동일 블록체인 장부를 분산  
관리하여 임의의 위변조/해킹 불가



#### Transparency

모든 거래기록에 대해 공개적으로  
조회 및 공동으로 Update



#### Decentralization

제3자의 검증 없이도 당사자 간  
거래에 대한 신뢰 확보



블록체인과 융합된 빅데이터는 **보안성**, **투명성**, **탈중앙화**가 적용되어 데이터의 가치를 제고

### 3. 데이터 기술유형

## RDB, NoSQL, 블록체인 등 크게 3가지 데이터 기술 특징점 비교

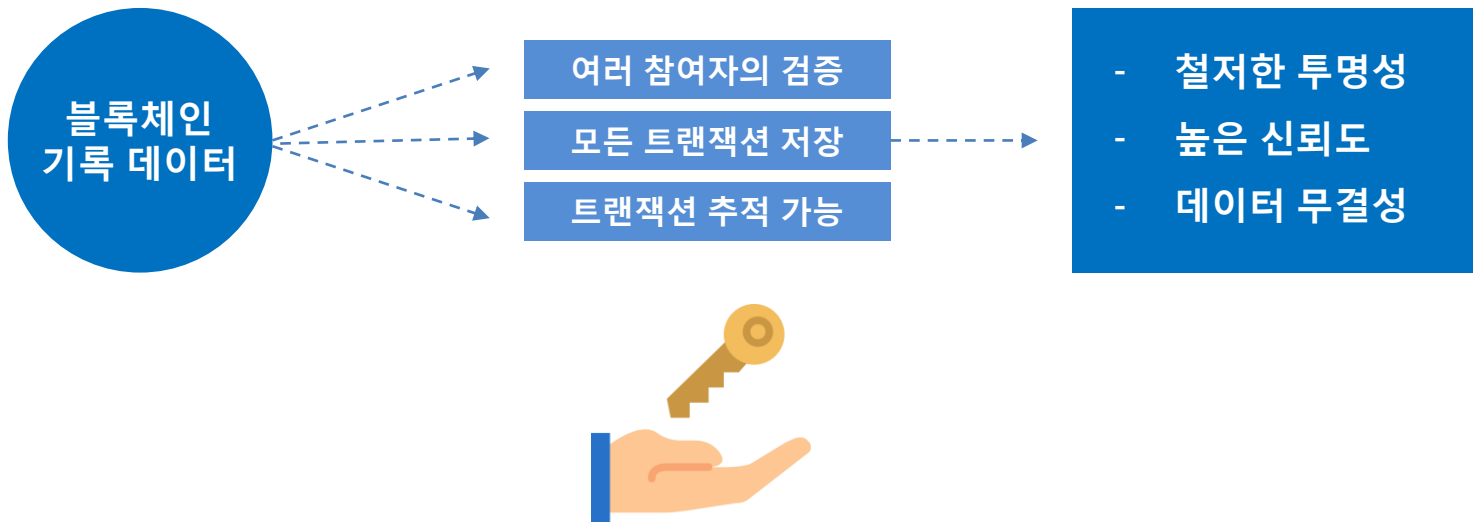
	RDB 관계형 데이터	NoSQL 빅데이터	블록체인 분산 데이터
이용 목적	Mission Critical한 데이터 및 정제된 데이터를 빠른 성능과 안정성을 위해 사용	각종 로그 데이터를 포함한 비정형화된 데이터를 처리하기 위해 사용 (데이터 사이즈가 얼마가 될지 예측이 불가능한 영역에 사용)	신뢰성과 보안이 강력히 요구되는 데이터를 처리하기 위해 사용
데이터 저장 위치	다양한 위치의 노드에 동일한 데이터가 분산되어 탈중앙화 형태로 저장	특정 서버에서 관리 (관리자 有)	
데이터 불가역성	기록이 완료된(합의된) 데이터는 변경 및 삭제 불가능 (create와 read만 가능)	데이터 변경 및 삭제 가능	
보안	데이터가 비가역적이므로 해킹으로부터 안전	다양한 해킹에 노출되어 있고 고가/고성능의 보안 솔루션 필요	

## 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

### 1. 데이터에 대한 강력한 믿음 '신뢰 보장(데이터 무결성)'

#### 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

## Trust





## 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

### 1. 데이터에 대한 강력한 믿음 '신뢰 보장(데이터 무결성)'

사례

SHA-256(Secure Hashing Algorithm 256)을 사용하여 데이터를 해싱 시

→ 데이터가 아무리 크든 작든 해시 값은 항상 256비트로 고정됨

→ 엄청난 양의 데이터와 트랜잭션을 처리할 때 효과적

→ 입력 데이터를 기억하는 대신, 해시 값을 기억하고 추적할 수 있음

Input	Output (Hash)
안녕하세요	3639EFCD08ABB273B1619E82E78C29A7DF02C1051B1820E99FC395DCAA3326B8
만나서 반갑습니다.	53A53FC9E2A03F9B6E66D84BA701574CD9CF5F01FB498C41731881BCDC68A7C8

#### 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

### 1. 데이터에 대한 강력한 믿음 '신뢰 보장(데이터 무결성)'

사례

SHA-256에서 입력 데이터를 조금만 바꿔도

→ 해시 값은 기존 데이터의 해시와 비교하여 연관성을 찾기 어려울 정도로 변함

→ 이처럼 데이터 위변조는 사실상 어려워 데이터 무결성을 보장

Input	Output (Hash)
안녕하세요	3639EFCD08ABB273B1619E82E78C29A7DF02C1051B1820E99FC395DCAA3326B8
안녕하세요.	2E99758548972A8E8822AD47FA1017FF72F06F3FF6A016851F45C398732BC50C25

#### 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

### 2. 데이터 해킹 No '악성 행위 방지'

## Fraud-Protection



#### 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

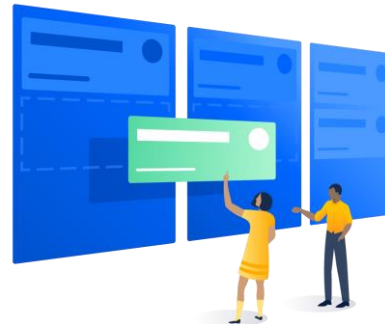
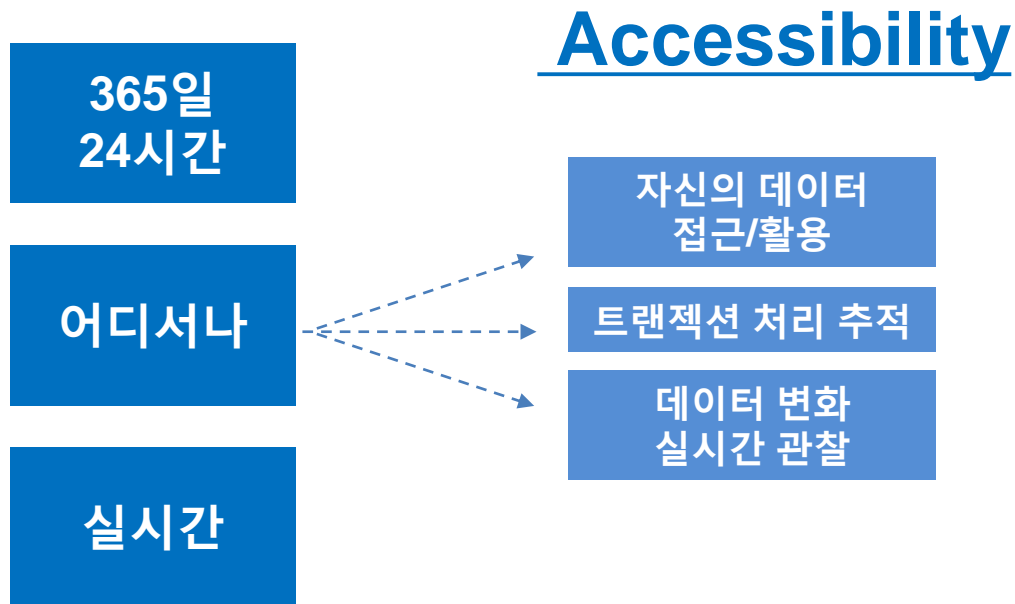
### 3. 정확한 데이터 분석 '예측 분석'

## Accuracy



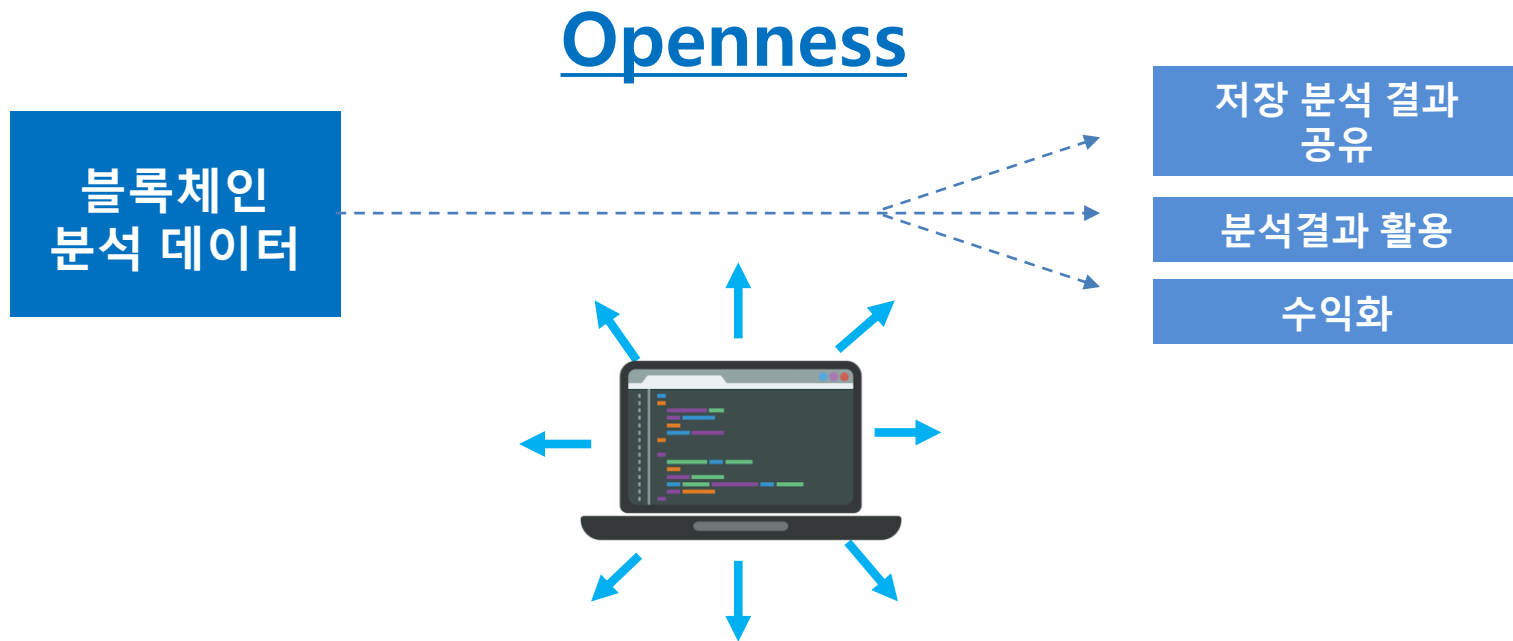
## 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

### 4. 데이터 즉시성 강화 '실시간 데이터 접근'



#### 4. 데이터 산업 적용을 위한 5가지 블록체인 특징

### 5. 모두의 데이터 '데이터 공유 관리'



## 5. 데이터 산업 적용 가능 블록체인 유스케이스

### 1. 개인정보 비식별화(De-Identification) > 개념

데이터 자체만으로  
개인을 식별할 수  
있는 정보는 원천  
삭제함

삭제 후 남아있는  
정보는 추가 가공  
(삭제, 변환 등) 후  
제공받는 자가  
보유한 정보와 공개  
정보를 결합하여  
누구도 정보 소유자  
개인을 식별할 수  
없도록 조치함

#### 가명처리

개인정보 중 주요 식별요소를 다른 값으로 대체하여 개인식별을  
곤란하게 함

홍길동, 35세, 서울거주, 한국대 재학

→ 임꺽정, 30대 서울 거주, 국제대 재학

#### 총계처리 또는 평균값 대체

데이터의 총합 값을 보임으로서 개별 데이터의 값을 보이지 않도록  
함

임꺽정 180cm, 홍길동 170cm, 이공취 160cm,

→ 물리학과 학생 키 합 : 510cm, 평균키 170cm

#### 데이터 값(가치) 삭제

데이터 공유 개방 목적에 따라 데이터 셋에 구성된 값 중에 필요없는  
값 또는 개인식별에 중요한 값을 삭제

홍길동, 35세, 서울거주, 한국대 졸업 → 35세, 서울 거주

주민등록번호 901206-1234567 → 90년대 생, 남자

#### 데이터 마스킹

공개된 정보 등과 결합하여 개인을 식별하는데 기여할 확률이 높은  
주요 개인식별자가 보이지 않도록 처리하여 개인을 식별하지  
못하도록 함

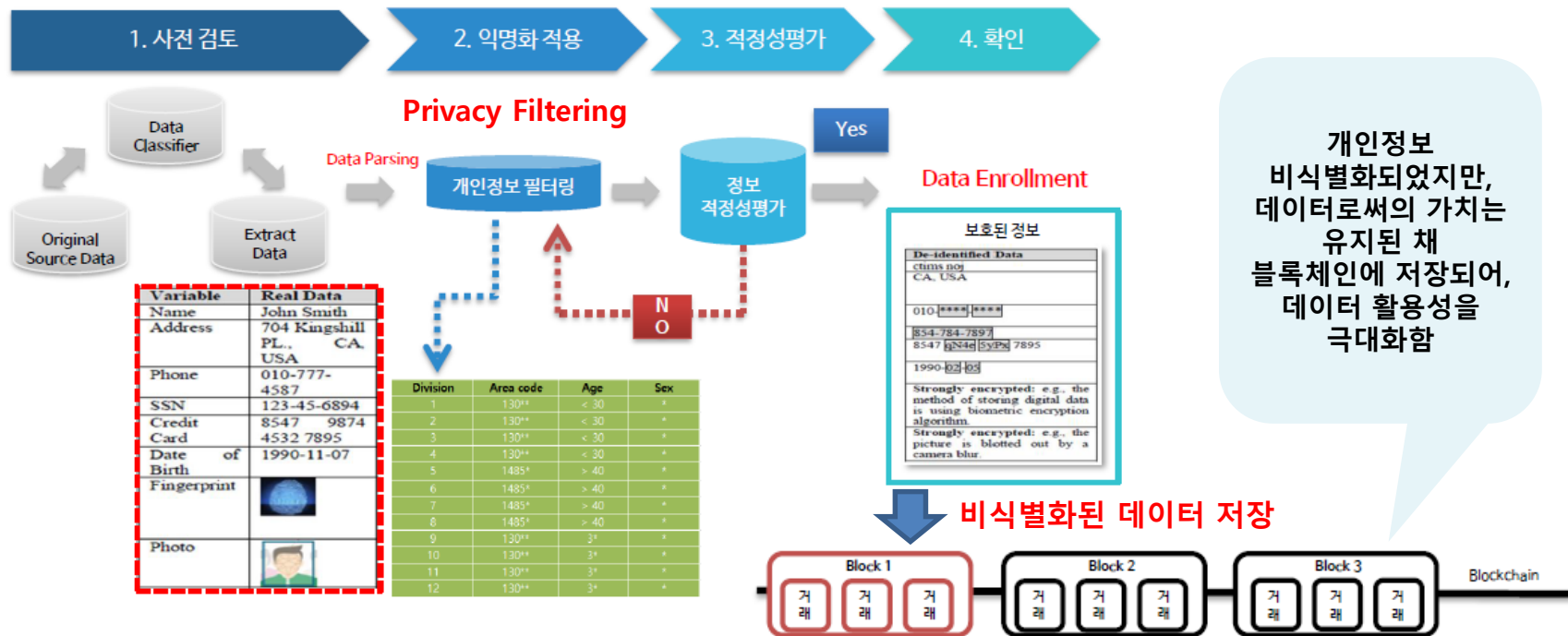
홍길동, 35세, 서울거주, 한국대 재학

→ 홍\*\*, 35세, 서울 거주, \*\*대학 재학

(“한국정보화진흥원” 자료 인용)

## 5. 데이터 산업 적용 가능 블록체인 유스케이스

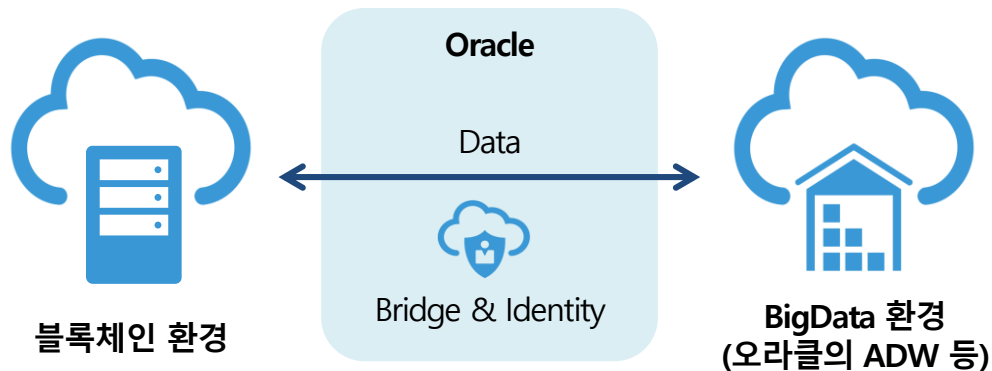
## 1. 개인정보 비식별화(De-Identification) > 처리방안





## 5. 데이터 산업 적용 가능 블록체인 유스케이스

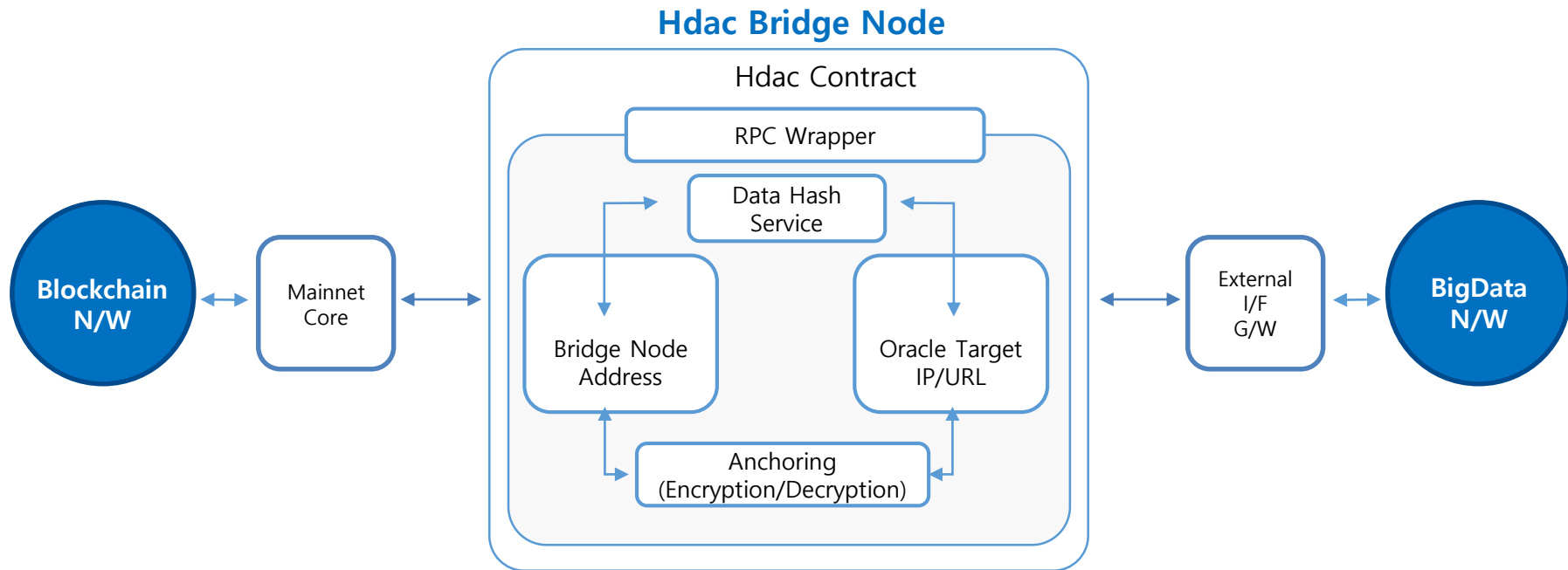
### 2. 자율 데이터 운영관리 (Autonomous Data Management) 강화



#### ※ 블록체인 오라클 문제란?

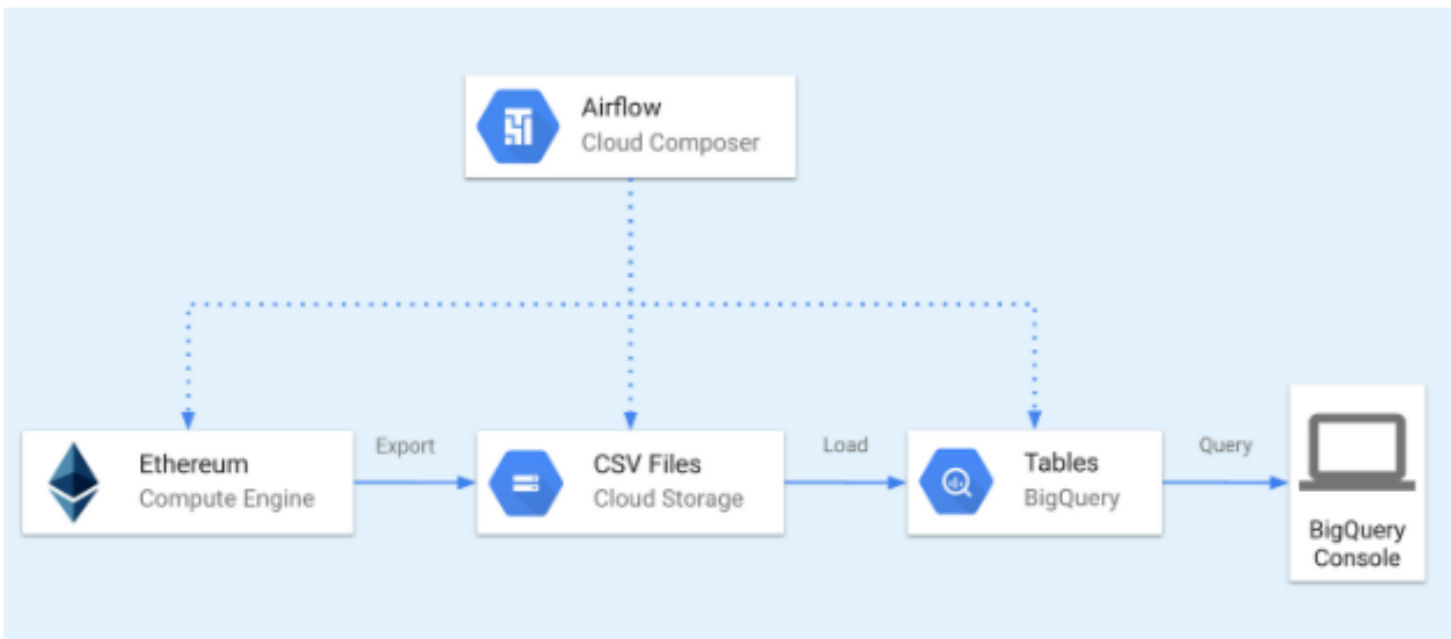
블록체인 외부에 있는 데이터를 블록체인 안으로 들여오고, 블록체인의 데이터를 외부로 내보내는 역할  
⇒ 블록체인 생태계를 **외부와 연결 시켜주는 하나의 다리(Bridge)** 역할을 하는 것을 오라클(Oracle)이라고 함

### 2. 자율 데이터 운영관리 (Autonomous Data Management) 강화



## 6. 별첨

# 이더리움 데이터 분석을 위한 Google BigQuery 아키텍처



## 6. 별첨

# Google BigQuery의 블록체인 데이터 분석 예시

### (사례) Popular Smart Contracts Event Logs

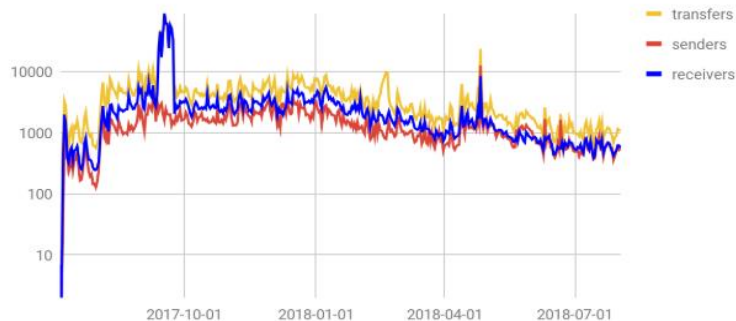
Top 10 이더리움(ERC-721) 컨트랙트 분석결과 최상단에 있는 0x06012c8 .... CryptoKitties 게임의 주소로 생일이벤트와 관련하여 가장 많은 트랜잭션을 확인.

Row	contract_address	tx_count
1	0x06012c8cf97bead5deae237070f9587f8e7a266d	2307079
2	0xd2f81cd7a20d60c0d558496c7169a20968389b40	53728
3	0x7fcd2a1e52f10c28cb7732f46393e297ecadda1	48556
4	0x2402b6fa4785866e8009f90da83d3bef27bb2ed3	45701
5	0xa98ad92a642570b83b369c4eb70efefe638bc895	39357
6	0x1a94fce7ef36bc90959e206ba569a12afbc91ca1	39185
7	0xbb5ed1edeb5149af3ab43ea9c7a6963b3c1374f7	33572
8	0xf7a6e15dfd5cdd9ef12711bd757a9b6021abf643	29291
9	0xe58af3dcbe423c1e330f9d6e61cd47cb530239ac	27370
10	0xd73be539d6b2076bab83ca6ba62dfe189abc6bbe	17524

### (사례) Transaction Volumes and Transaction Networks

가장 인기 있는 이더리움 토큰 중 하나인 OmiseGO의 일간 트랜잭션을 아래와 같이 비주얼화하여 확인 가능

OmiseGO Daily Transfers



## 6. 별첨

# 기타 블록체인 분석 플랫폼 소개 (LYZE, ChainAnalysis)

### LYZE

객관화된 정략적 수치 제공을 통해 암호화폐 투자에 활용할 수 있도록 제공하는 플랫폼. BTC, ETH, ERC-20 상위 42개 등의 On-chain 데이터를 수집

#### Key Features

##### 온체인 데이터 검색

BTC, ETH 및 LYZE에서 선정한 ERC-20 토큰의 온체인 데이터의 현황과 이에 대한 LYZE의 평가 내용을 제공

##### 온체인 활동 데이터 비교 검토

BTC, ETH 및 LYZE에서 선정한 ERC-20 토큰들을 카테고리 별, 온체인 데이터별 분류 및 비교

##### 일일 하이라이트 및 코인 추천

자체적으로 개발한 '온체인 데이터 활동과 가격의 상관관계의 기준'에 따라 매일 좋은 성적의 코인을 추천

#### On-Chain Data Types

- 신규지갑 수  
새로 생성된 토큰 보유 지갑 수
- 거래수  
암호화폐 전송 횟수
- 거래량  
암호화폐 전송 수량
- 온체인활동 종합지수  
신규지갑수, 활동지갑수, 총 누적 지갑수, 지갑별 자산, 지갑별 평가 손익 추정치 등을 취합하여 가중 평균화한 지수

### ChainAnalysis

자동화된 블록체인 데이터 분석을 제공하여 자금 세탁 등 암호화폐의 부정한 사용을 탐지 가능하게 해주는 솔루션. Bitcoin Cash, Ether, Litecoin과 주요 암호화폐를 지원

#### Key Features

##### Real-time transaction screening

대량의 활동을 지속적으로 관찰하여 고위험 트랜잭션을 식별

##### Case management capabilities

유저의 케이스 리뷰를 확인할 수 있고 AML 리포팅을 확인

##### Enhanced due diligence

개별 사용자 수준까지, 특정 거래에 대한 자금 흐름 추적

##### KYT for Token Issuers

토큰 발행자에게 불법 활동에 노출될 위험과 토큰 거래의 위험성에 대한 위험 평가 정보를 제공



---

**Thank You**

---