

# 고확장성 블록체인 저장 기술 설계

김창수, 이명철  
한국전자통신연구원 스마트데이터연구소  
[cskim7@etri.re.kr](mailto:cskim7@etri.re.kr), [mclee@etri.re.kr](mailto:mclee@etri.re.kr)

## Design of Scalable Blockchain Storage Technology

Changsoo Kim, Myungcheol Lee  
Smart Data Research Section, Electronics and Telecommunications Research Institute

### 요 약

블록체인은 블록체인 네트워크에 참여하는 모든 노드에 동일한 원장을 중복 저장함으로써 참여자 모두에게 데이터가 공개되며, 이를 통해 데이터의 투명성을 확보하고 참여자 다수의 합의를 통해 데이터의 신뢰성을 보장하는 기술이다. 현재의 블록체인 시스템은 이러한 데이터 소유의 탈중앙화를 통해 데이터의 신뢰성을 확보할 수 있게 되었으나, 합의, 저장, 분석 등 여러 방면에서 확장성 한계로 인해 주로 가상 자산의 거래에만 활용되고 있다. 본 논문은 다양한 일반 산업 및 비즈니스에서의 블록체인 기술의 활용성을 증대하기 위해 요구되는 데이터의 다양성, 대규모 용량, 대용량 데이터, 트랜잭션 처리 속도 등 저장 확장성 문제를 해결하기 위한 고확장성 블록체인 저장 기술을 제안하고 설계 개념 및 특징, 그리고 고확장성 저장 기술을 제공하기 위한 시스템 SW 구조를 제시한다.

### 1. 서론

최근 은행과 같은 금융 플랫폼, 구글, 네이버 등의 거대 데이터 플랫폼, 정부 기관 등과 같이 신뢰할 수 있는 제3자에 의한 개인 데이터 및 정보의 소유 집중화 문제를 해결하기 위한 방안으로서, 상호 신뢰하지 않는 다수의 참여자들 간에 분산된 장부들을 유지하고 이를 통해 데이터의 무결성 및 신뢰성을 확보하는 블록체인 기술이 주목받고 있다. 최근 3~4년간 비트코인, 이더리움 등의 가치가 급등하면서 가상 자산의 거래도 급증하고, 블록체인 기술에 대한 관심도 급격히 증가하고 있다.

기존 가상 자산 중심의 블록체인 기술은 점차 IBM의 Food Trust 솔루션[1]과 같은 일반 산업 및 비즈니스를 위한 기술로 발전하고 있다. MarketsandMarkets의 보고서[2]에 따르면, 비즈니스 프로세스의 단순화, 비즈니스 투명성과 불변성, 운용 비용 절감 등에 대한 수요 증가로 물류 및 공급망 관리 등에서 블록체인 기술에 대한 대중화가 시장을 이끌 것으로 예측하고 있다.

그러나, 보고서의 예측과 달리 여러 분야에서 응용 서비스 개발 및 사업화를 시도하고 있으나, 합의, 저장, 분석 등 블록체인이 가진 확장성 한계로 인해 시범 서비스 단계에 그치고, 여전히 가상 자산 위주

로 활용되고 있는 것이 현실이다.

향후 블록체인 기술이 가상 자산을 포함한 금융 분야에 한정되지 않고 다양한 산업 분야에서 활용성을 확대하기 위해서는 현재 블록체인 플랫폼이 가지는 이러한 제약 요소들을 해결하고 다양한 비즈니스 활용 상황에서 필요로 하는 기술의 확보가 필요하다.

따라서, 본 논문에서는 현재의 블록체인 시스템이 가지는 제약 요소들 중 저장 확장성 문제를 해결하고 다양한 산업 및 비즈니스 분야에서 활용성을 확대할 수 있는 고확장성 블록체인 저장 기술을 제시한다. 또한 제안하는 기술의 설계 개념, 특징 및 다양한 산업에서의 블록체인 서비스 제공을 위한 시스템 구조를 설명한다.

### 2. 고확장성 블록체인 저장 기술의 설계 이슈

향후 가까운 시일 내, 가상 자산을 포함한 금융 분야에서의 활용을 넘어 일반 산업에서의 활용성을 확대하기 위해 해결해야 하는 걸림돌로서, 현재의 블록체인 시스템은 여러 제약 사항이 존재한다. 현재의 블록체인 시스템은 자산의 전송 등에 특화되어 발전되어 왔기에 각 산업 분야에서 활용되는 다양한 종류의 대규모/대용량 데이터 저장에 한계를 갖는다. 특히나 최근 블록체인 시스템의 활용성이 급격히 증

가하면서 분산 원장의 크기가 급격히 증가하고 있고 머지않아 보통 수준의 서버에서 저장 및 처리가 불가능할 것으로 파악되고 있다.

따라서 점차 커지는 분산 원장을 효율적으로 분산 노드에서 관리하기 위한 기술 개발이 요구된다. 또한 주소-잔액 기준의 단순한 키-밸류 형태의 상태 데이터 관리를 위해 범용의 키-밸류 DBMS를 활용하는 방식을 탈피하여, 다양한 종류의 산업 데이터를 저장할 수 있고, 블록체인 특성을 고려하여 고속 처리할 수 있는 기술이 요구된다. 이와 더불어 기존 외부 저장소에 저장하던 대용량 데이터를 안전하게 블록체인 네트워크 내에서 관리할 수 있는 기술 역시 요구된다.

이러한 블록체인 저장 기술의 제약사항을 정리하면 표 1과 같다.

(표 1) 기존 블록체인 플랫폼의 저장 기술 한계

범주	제약사항
대규모 원장	- 중복 저장에 따른 저장공간 낭비 - 데이터 소유 집중화 유발
상태 데이터	- 범용 K-V DBMS - 저장공간 낭비 - 불필요한 연산 및 처리 속도 저하
대용량 데이터	- 외부 저장소 활용 - 가용성, 무결성, 소유권 보장 어려움

위에서 제시한 블록체인 저장 기술의 제약 사항들을 해결하기 위한 관련 연구가 국내외에서 일부 진행되고 있다.

규모가 확대되고 있는 대규모 원장 문제를 해결하기 위해 이더리움은 샤딩(Sharding)[3], 플라즈마 기술을 도입했다. 이러한 기술들은 블록체인 네트워크를 수평 분할하여 원장을 관리하거나, 체인을 메인 체인과 자식 체인으로 계층적 분할하여 관리하는 방식을 활용한다. 그러나 수평 분할된 체인은 비잔틴 공격에 취약성을 보이며, 계층적 분할은 전체적인 블록체인 데이터의 추적이나 분석이 어렵고 특정 자식 체인의 규모가 클 경우 기존의 취약성을 그대로 나타내는 문제가 있다.

상태 데이터 관리의 경우, 기존 키-밸류 저장소의 특정 알고리즘 성능을 개선[4]하거나, 랜덤 I/O 성능을 보완하기 위해 SSD를 활용하는 방법에 대한 연구[5] 위주로 진행되었으나, 트랜잭션 데이터의 용량이 커지면서 상태 데이터 용량도 급증하고 있으며, 블록체인에서의 상태 데이터 접근 워크로드를 고려하는 등의 블록체인 특성을 고려한 기술 연구는 미

흡한 실정이다.

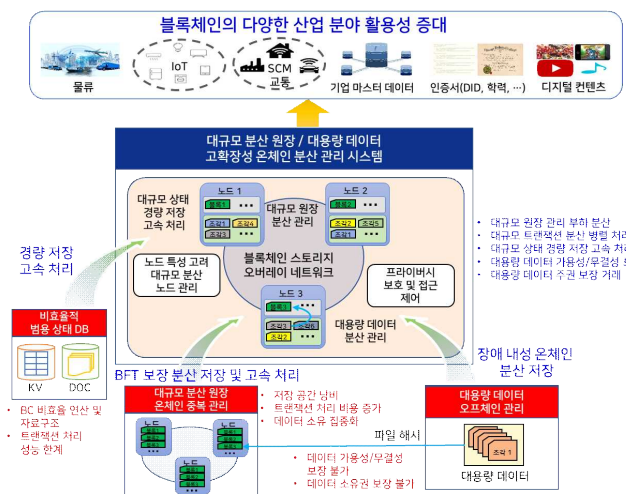
블록체인을 활용하면서 대용량 데이터를 지원하기 위한 방법으로 주로 IPFS[6]와 같은 별도의 대용량 데이터 스토리지 시스템에 원본 데이터를 저장하고, 이에 대한 해시값을 포함하는 메타데이터만 블록체인에 저장하는 방식을 활용하고 있다. 그러나, IPFS는 기본적으로 데이터 가용성을 보장하지 않고, 또한 블록체인과 별도의 데이터 관리 주체를 사용하면서, 대용량 데이터에 대한 가용성, 무결성 및 소유권 보장이 어려운 문제가 발생한다.

### 3. 고확장성 블록체인 저장 기술 설계

앞서 살펴본 현재 블록체인 시스템의 제약 사항과 함께, 블록체인 시스템이 향후 다양한 산업에서 활용될 상황을 고려하여 목표 시스템에 요구되는 요구사항을 분석할 필요가 있다. 요구사항을 위해 우리는 사용자 요구사항 및 시스템 요구사항을 범주별로 구분하여 분석하였다[7]. 요구사항의 범주로는 사용자 관리 및 접근제어, 대규모 분산 노드 관리, 대규모 트랜잭션 데이터 분산 저장, 대규모 상태 데이터 고성능 저장, 대용량 데이터 분산 저장이 있으며, 사용자 요구사항 40개, 시스템 요구사항 47개를 작성하였다.

설계 이슈 및 요구사항을 기반으로 설계를 진행하면서 고려한 기본적인 설계 개념은 확장성, 경량화, 고성능, BFT(비잔틴 장애 내성) 보장 등이다.

그림 1은 이러한 설계 개념을 지원하기 위한 고확장성 블록체인 저장 기술의 개념을 나타낸다.



(그림 1) 고확장성 블록체인 저장 기술 개념

그림 2는 설계 이슈와 요구사항 분석을 바탕으로

설계한 고확장성 블록체인 저장 시스템의 SW 구조를 나타내며 다음과 같은 요소들로 구성된다.

**서버 HW 플랫폼:** 서버 HW 플랫폼은 하나 이상의 프로세서, 메모리, 마더보드 등을 구비한 일반적인 HW 시스템이며 서버 HW 플랫폼 상에 고확장성 블록체인 저장 기술을 포함하는 블록체인 시스템이 동작한다.

**CSD 저장장치:** CSD(Computational Storage Drive) 저장장치는 PCI 버스에 연결된 저장장치로 FPGA나 ARM 코어 등 연산 가속 장치를 내장하고 내부 메모리를 가지고 있어 저장된 데이터의 인스토키지 처리(연산)가 가능한 장치이다. 이를 통해 SSD 내에 저장된 데이터를 서버 HW 플랫폼의 메모리에 올려서 CPU로 처리하지 않고, 저장장치 내에서 바로 처리하여 그 결과만을 서버 HW 플랫폼으로 전송하여, 데이터 처리 성능을 고속화하는 역할을 한다[8].

는 블록체인 네트워크를 구성하는 각 노드의 특성을 수집하고 이를 관리하며, 노드의 특성에 따라 데이터를 동적 저장하는 역할을 수행한다. 또한 노드의 상황이나 특성 변경에 따라 노드 간 부하 분산을 수행하고 노드 간 동적 라우팅을 관리한다.

**대규모 트랜잭션 데이터 고확장성 저장:** 1/3 노드까지 비잔틴 장애를 허용하고 노드 특성을 고려하면서 대규모 트랜잭션 데이터를 분산 저장하여 스토리지 낭비를 방지한다. 또한 인스토키지 기반 고성능 처리 기술을 활용하여 트랜잭션 데이터의 고속 처리를 지원한다. ECC(Error Correction Code), 레이저 코드 등 부호화 기술 기반의 블록 분할 및 분산 저장 기법을 활용하여 블록을 분산 저장한다.

**대규모 상태 데이터 고성능 저장:** 키/밸류, 키/문서 등 다양한 상태 데이터 종류를 지원할 수 있도록 상태 데이터 저장 기능을 지원하고, 블록체인에 최적화된 경량 상태 데이터의 관리 기능을 제공한다. 또한, 인스토키지 기반 고성능 처리 기술을 활용하여 상태 데이터의 고속 처리를 지원한다.

**대용량 데이터 온체인 분산 저장:** 다양한 산업 분야에서의 활용성을 증대시키기 위해 다양한 분야의 디지털 데이터(문서, 이미지, 동영상 등) 및 기업 데이터(DB 데이터)를 블록체인을 이용하여 관리할 수 있도록 온체인 내에 대용량 데이터를 장애 내성을 갖도록 분산 저장한다.

**인스토키지 기반 고성능 처리:** CSD 장치를 활용하여 저장장치 내에서 일부 연산을 수행함으로써 노드 내 서버 HW 플랫폼의 CPU의 활용성을 제고할 수 있도록 한다. 예로, K-V DBMS에 관리되는 상태 데이터의 컴팩션 작업을 저장장치 내에서 수행하거나, 부호화 기반 트랜잭션 데이터 분산 저장을 위한 인코딩/디코딩 알고리즘을 저장장치 내에서 수행함으로써 저장장치-시스템 메모리 간 불필요한 데이터 이동을 최소화하여 처리 성능을 향상한다.

**프라이버시 보호 및 접근 제어:** 블록체인 네트워크 내에 저장된 다양한 산업 분야의 대용량 데이터의 내용에 대해 인가된 사용자만이 접근할 수 있도록 하여 데이터의 프라이버시를 보호하고, 데이터 소유자의 정책에 따라 역할 기반 접근 및 다른 사용자로의 소유권 이전이 가능하도록 하여 데이터 소유자의 주권을 보장한다.

**대규모 대용량 블록체인 데이터 고확장성 분산 저장 API:** 블록체인 응용서비스에서 고확장성



(그림 2) 고확장성 블록체인 저장 시스템 SW 구조

**Linux OS:** 고확장성 블록체인 저장 기술을 포함하는 블록체인 시스템이 동작하기 위한 운영체제이다. 다수의 노드로 구성되는 블록체인 네트워크를 사용자가 쉽게 설치하고 관리할 수 있도록 Docker와 Kubernetes를 활용하여 블록체인 시스템을 Linux OS 상에서 컨테이너 기반으로 운영한다.

**시스템 관리:** 고확장성 블록체인 저장 기술을 탑재한 블록체인 시스템의 사용자를 관리하고, 블록체인 네트워크의 관리를 수행한다. 사용자의 역할, 권한을 관리하고 인증 등을 수행한다. 블록체인 네트워크 관리는 블록체인 네트워크의 생성, 채널 관리, 조직/피어 관리, 인증 서버 관리 등을 수행한다.

**대규모 분산 노드 관리:** 대규모 분산 노드 관리

블록체인 저장 시스템을 활용할 수 있도록 다양한 블록체인 데이터 저장, 접근 API를 제공한다.

#### 4. 결론

본 논문에서는 다양한 산업 분야에서의 블록체인 활용성 증대를 위한 고확장성 블록체인 저장 기술의 요구사항, 설계 개념 및 시스템 SW 구조를 제시하였다.

고확장성 블록체인 저장 시스템은 금융 분야를 넘어 다양한 산업과 비즈니스에서 블록체인 활용성을 증대하기 위해 요구되는 블록체인 저장 확장성 기술들을 제공하는 시스템이며, 확장성, 경량화, 고성능, BFT 보장 등의 설계 개념과 다양한 범주의 요구사항을 고려하여 설계를 진행하였다.

고확장성 블록체인 저장 기술은 대규모 트랜잭션 데이터와 상태 데이터의 고속 처리를 위해 CSD 장

치를 활용한 인스토키지 처리 기술을 채택하였고, 데이터 중복을 최소화 하면서 비잔틴 장애 내성을 유지할 수 있도록 부호화 기반 데이터 분할 및 분산 저장 방식을 채택하였다. 또한 대용량 데이터를 장애 내성을 갖도록 온체인 분산 저장하여 대용량 데이터의 무결성, 가용성 및 소유권을 보장할 수 있다.

향후 연구는 본 설계 내용을 기반으로 각 요소기술에 대한 상세 설계를 진행하고 시스템 프로토타입을 개발하여, 본 설계 내용에 대한 검증을 진행할 예정이다.

\* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로, 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00136, 다양한 산업 분야 활용성 증대를 위한 대규모/대용량 블록체인 데이터 고확장성 분산 저장 기술 개발).

#### 참고문헌

[1] IBM, “IBM Food Trust: 전 세계 식품 공급의 새로운 시대가 열립니다.”, <https://www.ibm.com/kr-ko/blockchain/solutions/food-trust> (2021.09.29. 방문)

[2] MarketsandMarkets, “Blockchain Market by Component (Platform and Services), Provider (Application, Middleware, and Infrastructure), Type (Private, Public, and Hybrid), Organization Size, Application Area (BFSI, Government, IT & Telecom), and Region - Global Forecast to 2025”, May 2020.

[3] 해시넷, “샤딩”, <http://wiki.hash.kr/index.php/샤딩> (2021.09.29. 방문)

[4] 김재은 외 3명, “다중레벨병합을 통한 LevelDB 성능 향상”, 한국정보과학회 학술발표논문집, 2018.

[5] B. Lepers, et al., “KVell: the design and implementation of a fast persistent key-value store”, SOS2019

[6] IPFS homepage, “<https://ipfs.io>” (2021.09.29. 방문)

[7] 김창수 외 7인, “다양한 산업 분야 활용성 증대를 위한 대규모/대용량 블록체인 데이터 고확장성 분산 저장 기술 요구사항정의서”, 한국전자통신연구원, 2021.

[8] Torabzadehkashi et al, “Computational storage: an efficient and scalable platform for big data and HPC applications”, Journal of Big Data, 2019.