

블록체인 분석 플랫폼을 위한 시계열 인덱스 설계 및 구현

원종호, 장미영, 설동명, 김지용

¹ 한국전자통신연구원 초지능창의연구소 지능정보연구본부 스마트데이터연구실

jhwon@etri.re.kr, myjang@etri.re.kr, dmsul@etri.re.kr, kjy@etri.re.kr

Design and Implementation of a Time-series Index for Blockchain Analysis Platform

Jongho Won, Mi-Young Jang, Dong-Myung Sul, Ji-Yong Kim
Smart Data Research Section,
Superintelligence Creative Research Laboratory,
Intelligence Information Research Division
Electronics and Telecommunications Research Institute (ETRI)

요 약

블록체인 분석 플랫폼은 블록체인에 저장된 데이터 기반의 다양한 산업분야 활용성 증대를 위하여 분산 블록체인 기반 대규모/대용량 데이터에 대한 고속 분석을 통하여 신뢰성이 보장되는 보안과 신뢰 기반의 데이터 서비스를 제공하기 위한 분석 플랫폼이다. 본 논문에서는 블록체인 분석 플랫폼에서 제공하는 데이터 분석 중 시계열 데이터에 대한 고성능의 분석을 제공하기 위한 시계열 데이터 인덱스의 설계와 구현에 대하여 기술한다.

1. 서론

블록체인은 블록체인 네트워크에 참여하는 모든 구성원(참여 노드)이 분산형 네트워크를 통해 블록체인에 저장되는 정보를 검증하고 저장/관리하여, 특정 구성원에 의한 임의적인 조작을 제한함으로써, 신뢰성이 보장되는 보안과 신뢰 기반의 데이터 서비스가 가능한 장점으로 인하여 블록체인 기반의 서비스에 대한 요구가 지속적으로 증가하고 있다. 이에 한국전자통신연구원에서는 다양한 산업분야 활용성 증대를 위한 분산 블록체인 기반 대규모/대용량 데이터 고속 분석을 제공하는 블록체인 분석 플랫폼을 개발하고 있다.

본 논문은 블록체인 분석 플랫폼에서 시계열 데이터에 대한 고성능의 분석을 제공하기 위한 시계열 데이터 인덱스의 설계와 구현에 대하여 기술한다.

본 논문의 구성은 다음과 같다. 2 장에서는 블록체인 분석 관련 연구를 소개하고, 3 장에서는 블록체인 분석 플랫폼의 구조에 대하여 기술한다. 4 장에서는 블록체인 분석 플랫폼의 시계열 데이터에 대한 시계열 인덱스 설계 및 구현에 대하여 기술하고, 마지막으로 5 장에서는 결론에 대하여 기술한다.

2. 관련 연구

블록체인 네트워크 기반으로 신뢰성이 보장되는 다양한 서비스를 제공하기 위하여, 블록체인 네트워크에 저장된 데이터를 신속하게 검색하여 제공할 필요가 있으나, 지속적으로 증가하는 블록체인에 저장된 데이터 속에서 단순한 트랜잭션 결과를 제외하고 유의미한 데이터의 검색은 제한되어 있다. 블록체인 데이터는 시간에 따라 지속적으로 증가하고, 대규모/대용량 블록체인 데이터의 검색과 분석의 필요성은 증가하고 있지만, 블록체인 데이터 분석을 위한 비효율적인 데이터 검색 성능으로 인하여 블록체인 데이터에 대한 분석 효율이 저하되고 있다.

이를 해결하기 위하여 블록체인 네트워크에 저장되는 데이터의 고성능 분석을 위한 다양한 연구와 시도가 제안되고 있으며, 크게 1) 블록체인 자체에 고성능 검색 기술의 제공하기 위한 온라인 블록체인 분석과 2) 블록체인과 별도의 검색 기능을 제공하는 오프라인 블록체인 분석 기술, 그리고 3) 온라인 블록체인 분석과 오프라인 블록체인 분석의 장점만을 활용하는 하이브리드 블록체인 분석으로 구분될 수 있다.

온라인 블록체인 분석은 주로 DBMS 기술을 활용

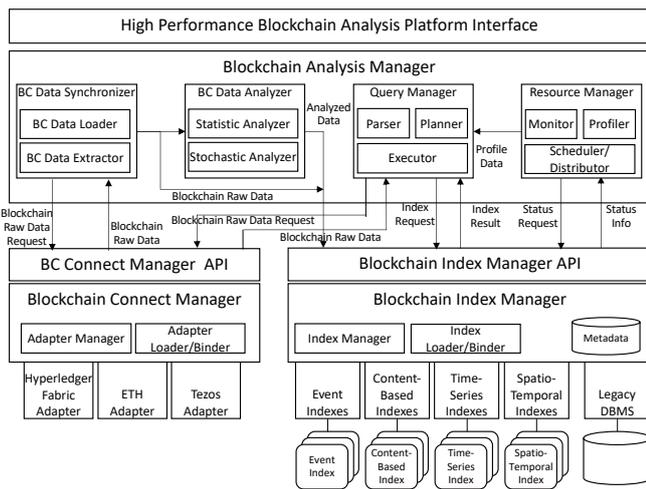
하여 블록체인 네트워크를 구현하는 것으로, 수십 년 간의 연구/개발을 통해 성숙된 데이터베이스 기술을 블록체인 네트워크에 활용함으로써 블록체인 네트워크의 안정성을 확보할 수 있는 장점을 가지고 있는 반면, DBMS 구현의 복잡성으로 인한 블록체인 네트워크 구현이 어렵고 블록체인 네트워크가 복잡해지는 단점을 갖는다[1].

오프라인 블록체인 분석은 블록체인 네트워크와 별도의 검색 환경을 구축하여 블록체인 데이터 검색 성능을 향상시키기 위한 시도로, 블록체인 네트워크에 저장되는 데이터를 별도의 DBMS 와 같은 검색 환경에 저장 관리하는 것으로, 고성능의 검색 기능을 제공하는 장점을 제공하는 반면, 블록체인 네트워크와의 일관성 보장이 어려운 단점을 갖는다[2,3].

하이브리드 블록체인 분석은 온라인 블록체인 분석과 오프라인 블록체인 분석의 단점을 극복하기 위하여, 검색의 대상이 되는 데이터는 블록체인 네트워크의 데이터를 사용하고, 관련 데이터에 대한 인덱스 등의 검색 기능만을 별도로 구축함으로써, 오프라인 블록체인 검색의 일관성 보장 문제를 해결하고, 온라인 블록체인 분석의 복잡성 문제를 해결하기 위하여 제안되고 있다[4,5].

3. 블록체인 분석 플랫폼

블록체인 네트워크에 저장되는 데이터의 고성능 분석을 제공하기 위한 하이브리드 블록체인 분석 방식의 블록체인 분석 플랫폼의 구조는 그림 1 과 같다.



(그림 1) 블록체인 분석 플랫폼 구조

블록체인 분석 플랫폼은 그림 1 과 같이 다양한 종류의 블록체인 네트워크를 연결하는 블록체인 연결 관리자, 블록체인 네트워크에 저장된 데이터에 대한 색인을 제공하는 블록체인 인덱스 관리자, 그리고 블록체인 네트워크에 저장된 데이터에 대한 분석을 제

공하는 블록체인 분석 관리자로 구성되며, 블록체인 분석 플랫폼을 사용하고 연동하기 위한 블록체인 분석 플랫폼 인터페이스를 제공한다.

블록체인 연결 관리자는 블록체인 분석 플랫폼의 독립적인 인터페이스를 통해 블록체인 분석 관리자와 인덱스 관리자가 요청하는 블록체인 데이터를 제공하는 기능을 제공한다. 블록체인 인덱스 관리자는 블록체인 데이터를 빠르게 검색하고 접근하기 위한 데이터를 색인하여 저장하고 관리하며, 시계열/시간간/이벤트 및 내용기반 색인 기능을 제공한다. 블록체인 분석 관리자는 블록체인 데이터와 인덱스 간의 데이터 동기화를 관리하고, 블록체인 분석 플랫폼에서 관리하는 데이터에 대한 질의 처리를 통한 블록체인 데이터 분석을 지원하며, 블록체인 분석을 위한 다양한 자원 모니터링을 통한 성능 향상을 지원한다.

4. 블록체인 분석 플랫폼을 위한 시계열 인덱스

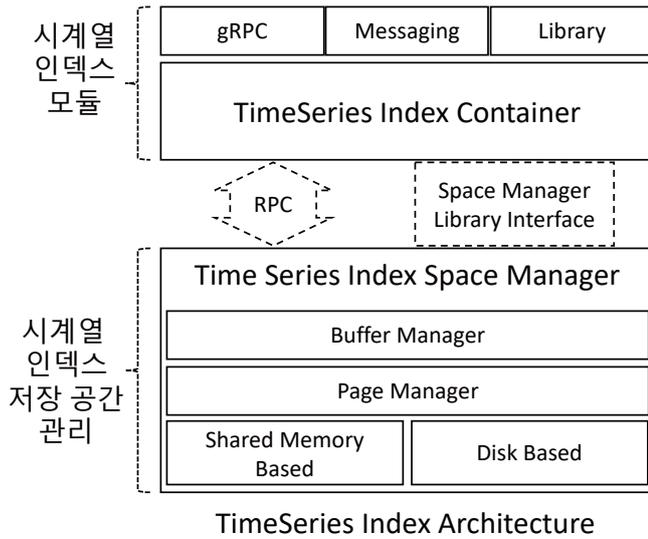
시계열 데이터는 시간의 흐름에 따라 관찰되는 시간을 기준으로 측정된 자료로, 연도별, 분기별, 월별, 일별, 시간별 등 시간의 경과(흐름)에 따라 순서대로 관측(측정)되는 자료로, 예를 들어 연간 국내 총 생산, 물가지수, 판매량, 종합주가지수, 강수량, 실험 및 관측 자료 등이 있다.

시계열 데이터 분석을 위한 블록체인 분석 플랫폼의 시계열 인덱스는 다음과 같은 특징을 갖는다.

- 색인을 위한 시계열 데이터의 시간 정보는 블록체인의 트랜잭션 내에 포함된 시간 정보를 대상으로 한다.
- 시계열 데이터는 동일한 시간 주기 별 지속적으로 발생한다.
- 시계열 인덱스에 추가되는 시계열 데이터의 시간 정보는 순서를 보장하지 않는다.
- 시계열 인덱스에 색인되는 특정 시점의 데이터 개수는 가변이다.
- 색인되는 데이터 크기는 고정이다 (블록체인 트랜잭션 식별자)
- 색인되는 데이터에 대한 변경(갱신) 또는 삭제는 발생하지 않는다.

블록체인 분석 플랫폼을 위한 시계열 인덱스의 특징을 고려한 시계열 인덱스 구조는 그림 2 와 같다.

블록체인 분석 플랫폼을 위한 시계열 인덱스는 인덱스를 저장하고 관리하기 위한 시계열 인덱스 저장 관리 모듈과 블록체인 분석 플랫폼에 저장되는 시계열 데이터를 색인하고 관리하기 위한 시계열 인덱스 모듈로 구성된다.



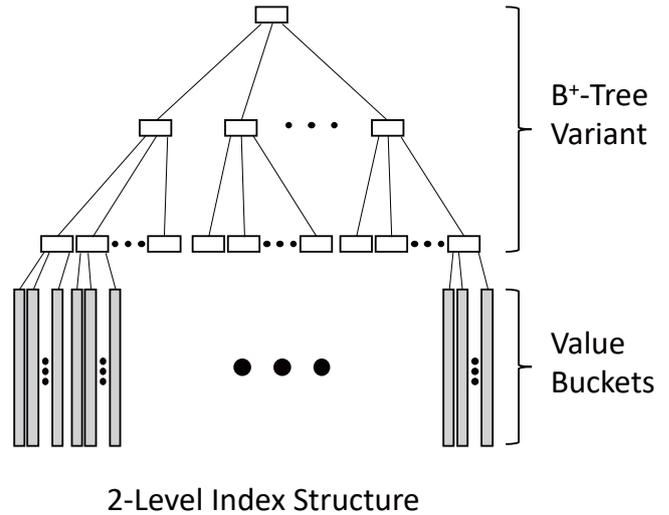
(그림 2) 블록체인 분석플랫폼을 위한 시계열 인덱스 구조

시계열 인덱스 저장 공간 관리 모듈은 지속적으로 증가하는 시계열 데이터의 특징을 반영할 수 있도록 볼륨 단위로 디스크 또는 메모리 저장 공간의 확장을 지원한다. 하나의 볼륨 크기는 운영체제가 허용하는 파일의 최대 크기까지 가능하며, 시계열 인덱스 저장 공간은 사전 정의된 최대 64 개의 볼륨으로 구성될 수 있고, 시계열 인덱스를 저장하고 관리하는 저장 공간은 사전 정의된 페이지 크기 (4KB) 단위로 관리된다. 저장공간과 메모리 사이의 가상화를 위한 페이지 단위 버퍼 관리 기능을 제공하며, 버퍼의 페이지 교체는 LRU (Least Recently Used, 가장 오랫동안 사용되지 않은 페이지 교체 정책을 사용한다.

블록체인 분석플랫폼의 시계열 인덱스를 쉽게 사용할 수 있도록, 시계열 인덱스는 C 라이브러리 인터페이스, TCP/IP 메시지 인터페이스와 gRPC 인터페이스와 같은 다양한 인터페이스를 제공한다.

시계열 인덱스는 시계열 인덱스 저장 공간의 효율성과 성능을 위해 시간 정보 색인을 위한 B⁺-트리와 특정 시점의 복수개의 데이터를 위한 버킷(bucket) 구조로 구성되어, 시간 정보를 데이터 정보를 별도로 구분하여 저장/관리하는 그림 3 과 같은 2-레벨 인덱스 구조를 갖는다.

시계열 인덱스에 지속적인 증가를 보장하지 않는 시간 정보를 색인하기 위해 시간 데이터는 일반적인 B⁺-트리 구조를 갖으며, 시간 범위 및 시간 윈도우 기반의 검색을 지원하기 위하여 B⁺-트리의 리프 노드는 양방향 연결 구조를 갖는다. 특정 시점에 대한 가변 개수의 데이터를 저장 관리하기 위해 데이터를 저장하는 버킷은 단방향 연결 구조로, 이를 통한 단일 시간 기반 검색을 효율적으로 처리할 수 있다.



(그림 3) 시계열 인덱스 구조

5. 결론

본 논문에서는 블록체인 분석 플랫폼을 위한 시계열 인덱스 설계 및 구현에 대하여 기술하였다. 시계열 데이터의 지속적으로 증가하는 시계열 데이터의 특징을 반영하는 인덱스 저장 공간을 설계하고 구현하였으며, 동일한 단위 시간에 대한 다수의 값(트랜잭션 해시 값)을 효율적으로 저장/관리하기 위한 2-레벨 인덱스 구조를 갖도록 하였다.

앞으로, 블록체인 분석 플랫폼의 시계열 인덱스에 대한 일관성 제공을 위한 트랜잭션 처리와 시계열 인덱스 성능 향상을 위한 효율적인 동시성 관련 연구 및 개발을 진행할 예정이다.

ACKNOWLEDGEMENTS

본 논문은 2023 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00180, 다양한 산업 분야 활성화 증대를 위한 분산 저장된 대규모 데이터 고속 분석 기술 개발)

참고문헌

- [1] S. Nathan, et all., "Blockchain Meets Database: Design and Implementation of a Blockchain Relational Database," IBM Research India, Proc. of VLDB Endowment, 12(11) July 2019
- [2] Conseil, <https://github.com/Cryptonomic/Conseil>
- [3] lib_indexer, <https://gitlab.com/nomadic-labs/tezos-indexer/-/tree/master/src>
- [4] TzScan indexer, <https://tzscan.io/>
- [5] Blockwatch, <https://tzstats.com/blog/next-gen-blockchain-indexing-for-tezos>