

블록체인 기반 데이터베이스 성능 분석 모델에 대한 제안

박제호^{*†}

^{**}단국대학교 소프트웨어학과

Proposal of Performance Analysis Model for Blockchain-based Database System

Je-Ho Park^{*†}

^{**}Dankook University, Dept. of Software Science

ABSTRACT

When blockchain technology, which shows various applicability, is utilized as a component of a database system, the characteristics of open verification and integrity/transparency of blockchain technology can bring new functionalities or enhanced results to the existing database system. However, when applying this blockchain technology to a database system, the cost versus expected effect in various performance perspectives must be evaluated. These costs include execution time and required storage space, and the performance of the converged system may vary in analysis method depending on the configuration method of the characteristics of the blockchain. This paper aims to propose an analysis model for the entire architecture by considering aspects that are not considered in the performance analysis models of database systems and the unique characteristics of blockchains. In doing so, we are trying to build a theoretical framework as an important conceptual technique that should be considered in the evaluation process of the performance results that can be obtained through the utilization of blockchain components in database systems. What we hope is that this work is expected to provide a useful foundation for researchers interested in the convergence of database systems and blockchain technology in order to construct a system with new future functionalities.

Key Words : Database System, Blockchain, Performance Evaluation, Modeling, Conceptual Technology

1. 서 론

일반적인 데이터베이스 시스템의 성능 분석은 기본적으로 트래잭션 수행에 필요한 처리속도 또는 응답속도(response time), 단위 시간에 처리할 수 있는 처리량(throughput), 자원 사용량(resource utilization), 잠금 및 동시성 관리(locking and concurrency management) 등의 성능 속성을 평가하고 고려하며, 확장성(scalability), 안정성(reliability), 오류율(error rate) 등의 속성도 부가적으로 고려를 하여야 한다[1].

이러한 데이터베이스 시스템의 성능 분석 모델의 요소 중 응답시간, 처리량, 자원 사용량, 동시성, 확장성, 오류율과 같은 성능 속성은 정량적인 측정을 통하여 직접적으로 확인할 수 있는 분석 요소이다. 반면에, 사용자 경험, 확장성, 안정성, 보안성(security) 실제적 측정에 의존하기보다는 개념적 수준이나 디자인 분석을 통한 속성도 존재를 한다. 특히, 확장성은 직접적인 분석인 가능하면서도 상황에 따른 변동에 따라 상황의존적 평가결과를 가져올 수 있는 이유로 예상 부하, 확장 방식, 사용되는 환경 등과 연관성을 주의해서 분석을 하여야 한다. 직접적 측정이 어려운 평가 요소 중 보안성은 데이터 기밀성, 무결성, 가용성 등

[†]E-mail: dk_jhpark@dankook.ac.kr

을 보호하는 능력을 의미하며, 암호화, 접근제어, 감사 추적 등의 기술적 조직적 요소로 평가를 진행한다.

이러한 데이터베이스 시스템의 평가 분석 요소들은 실제 데이터 관리를 하기 위한 처리 과정과 저장형태에 따른 분석 측면을 고려하여 성능분석 모델을 구성한다. 또한, 데이터베이스 시스템의 구조는 처리 및 저장 노드가 단순하게 하나의 노드로 구성되는 중앙집중식과 다수의 노드에 처리와 저장이 분산되는 형태로 분류할 수 있다. 이 두 형태의 장단점에 대해서는 이미 많은 연구가 진행되었다[2]. 1974년 이후 데이터베이스의 소개 이후 광대한 발전을 하여 왔으며, 그 과정을 통해서 기능성이 확장되고 개선되었으며, 아울러 필요한 구조 및 성능 분석 모델도 많은 변화가 있었다. 본 논문에서는 성능 분석에 필요한 요소를 포함하는 개념적 데이터베이스 시스템을 대상으로 논의를 진행한다.

2008년 비트코인과 함께 구성 컴포넌트의 개념으로 알려진 블록체인은 2014년 이더리움에 공개 시스템으로 채용되었으며, 2015년 엔터프라이즈 시스템에 사용되어 개인과 조직 간의 프라이빗 블록체인 네트워크 기술로 발전하였으며[3, 4], 다양한 분야에 응용되고 있다[5,6]. 일반적인 블록체인은 구성요소로서 단위적 데이터 레코드를 저장하는 블록을 사용하며, 이러한 블록을 상위 개념의 집합적 저장소를 체인이라고 하는 저장구조를 가지고 있다. 각 블록은 사용자 데이터 뿐만 아니라 보안을 위한 해쉬값을 이용하며, 해쉬값은 체인을 구성하기 위하여 구조적으로 관계된다. 이러한 저장소를 기반으로 하여 블록체인은 P2P 네트워크의 각 노드의 저장구조 역할을 담당하게 되며, 각 블록체인 노드 사이의 협업을 통해 공개 검증 기능과 무결성/투명성 보장이라는 블록체인 특유의 기능성을 가능하게 한다. 여기서 무결성은 모든 사용자는 검색하려고 하는 데이터가 기록된 시점 이후로 변경 또는 손상되지 않았음을 확인할 수 있음을 의미하고, 투명성은 모든 사용자가 과거 특정 시점 이후로 블록체인 체제에 어떤 데이터가 추가되면서 데이터 양이 증가 되어 왔는지를 확인할 수 능력을 의미한다[7-10].

본 논문에서 논의하고자 하는 성능 분석 모델의 구성은 블록체인의 속성 중에서 간과될 수 있는 측면을 성능 분석 모델에 포함하여 실제적으로 블록체인이 데이터베이스 시스템에 하부 컴포넌트로 포함되는 경우 수행 시간과 저장 필요 공간에 대한 비용 대비 효과에 대한 예측을 가능하게 할 수 있는 분석 모델을 구성하고, 이러한 모델을 기반으로 하여 전체 데이터베이스 아키텍처의 디자인을 결정하는 과정에서 간과할 수 있는 정서적 측면으로서 성능 분석 결과를 제공할 수 있도록 하기 위함이다.

본 논문은 2장에서 데이터베이스 시스템의 구성을 개

념적으로 포괄하고, 블록체인의 필요 구성 요소에 대한 개괄을 정리한다. 또한, 두 시스템을 융합하여 하나의 시스템으로 연동하는 경우의 성능적 요소를 기반으로 하여 성능에 대한 개념 모델을 제공한다. 마지막으로 3장에서 전체 내용에 대한 결론을 제안한다.

2. 본 론

제안하는 성능 분석 모델을 논의하기 이전에 먼저 필요한 데이터베이스 환경을 논의하고, 다음에 블록체인의 전체적인 개념을 정리하고 이를 데이터베이스 시스템과 융합하는 모델을 논의한다. 현대 데이터베이스 시스템은 기능, 구조, 목적 등에 따라 다양한 구조 및 운영 시스템으로 구성되지만, 본 논문에서는 가장 기본적인 모델을 기반으로 하여 성능 분석 모델을 구성하고자 한다. 기본적인 개념적 데이터베이스 시스템 환경은 크게 세 개의 개념적 컴포넌트로 구성된다: 사용자 계층, 데이터베이스 운영 계층, 데이터베이스 저장 계층. 이 구성은 그림 1에 예시되어 있다.

2.1 개념적 데이터베이스 성능분석 모델

현대 데이터베이스 시스템은 기능, 구조, 목적 등에 따라 다양한 구조 및 운영 시스템으로 구성되지만, 본 논문에서는 가장 기본적인 모델을 기반으로 하여 성능 분석 모델을 구성하고자 한다. 기본적인 개념적 데이터베이스 시스템 환경은 크게 세 개의 개념적 컴포넌트로 구성된다: 사용자 계층, 데이터베이스 운영 계층, 데이터베이스 저장 계층. 이 구성은 Fig.1에 예시되어 있다[11].

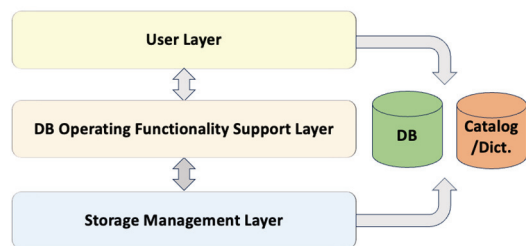


Fig. 1. General Database Architecture with 3-tier Layers.

Fig. 1에서 사용자 계층(user layer)는 사용자 측면에서 데이터베이스 시스템을 정의하고 활용할 수 있는 기능을 지원한다. 사용자 범주에는 데이터베이스 시스템 환경을 관리하는 권한을 가지는 관리자(DBA: database administrator)를 포함한다. 그림에서 사용자 계층이 오른쪽에 그려진 저장 계층과의 인터페이스를 가지는 이유는 데이터 모델

링 관련해서 디자인적 요소가 포함되기 때문이다. 데이터베이스 운영 지원 계층 (DB operating functionality support layer)는 데이터베이스 시스템이 지원해야 하는 기능을 총괄적으로 포함하는 범주로 데이터의 입출력을 포함한 사용자와의 상호작용 및 하부 컴포넌트 관리에 대한 요구 관리를 담당한다. 이 두 개의 계층은 개념적으로 소프트웨어의 영역이므로 분석 성능 지표는 요구에 대한 결과 반환까지의 응답시간에 비중을 둘 수 있다. 저장소 관리 계층 (storage management layer)는 데이터베이스가 포함하는 사용자 및 시스템 정보를 저장하는 기능을 지원하며 이는 소프트웨어 측면과 저장 하드웨어의 융합된 측면을 가지고 있다. 여기서, DBMS(database management system)는 트랜잭션의 안전한 수행을 보장하기 위해 ACID (원자성, 일관성, 무결성, 지속성) 속성을 이용한다. 이들 중에 무결성과 지속성을 지원하기 위하여 DBMS 저장 계층은 데이터베이스 로그(DB log) 또는 바이너리 로그를 유지하며, 이 로그는 로그 일련번호, 마지막 레코드에 대한 참조, 트랜잭션 ID, 레코드 유형 등을 포함한 정보를 축적한다[11-13]. 데이터베이스 시스템과 블록체인을 융합할 때 개선될 수 있는 여러가지 장점 중에 이 논문에서는 데이터베이스의 ACID 능력에 초점을 맞추고 특히 로그를 이용한 데이터베이스 시스템의 무결성 유지와 지속성 확보 측면에 대해 논의를 하고자 한다. 사용자 요청에 대한 성능 분석은 크게 다음과 같이 네 가지 성능 분석 속성으로 구성된다.

- T_{Exec} = 사용자 요청 처리에 필요한 과정을 수행하기 위해 소요되는 시간
- $T_{Storage}$ = 사용자 요청 처리에 필요한 데이터베이스 저장 계층과의 인터페이스와 관련 요청을 만족하기 위해 소요되는 시간
- T_{Backup} = 사용자 요청 처리 외에 백업에 필요한 과정을 수행하기 위해 소요되는 시간
- $T_{BKStorageProc}$ = 백업에 필요한 과정을 수행하기 위해 소요되는 시간
- $T_{BKNetwork}$ = 백업을 위해 네트워크를 통한 절차가 있는 경우, 네트워크 상에서 소요되는 시간

따라서, 하나의 사용자 요청에 대한 응답시간은 위의 네 가지 요소를 모두 포함한다는 정의에 대해서 일반적으로 다음과 같이 표현된다.

$$T_{UserReq} = T_{Exec} + T_{Storage} + T_{Backup} \quad (1)$$

$$T_{Backup} = T_{BKStorageProc} + T_{BKNetwork} \quad (2)$$

본 논문의 논의에서 주체적인 이슈는 데이터베이스 시스템과의 융합이 일반적으로 장점이 확정될 수 없기에

이를 분석할 수 있는 성능 분석 모델을 제안하고자 한다. 전술한 바와 같이 본 논문은 주어진 환경이 데이터베이스 로그를 포함하는 구조라는 전제 하에서 논의를 하고자 하는 것이므로, 식 1에서 볼 수 있는 성능 요소 중 블록체인을 데이터베이스 시스템에 융합하여 변화가 일어날 가능성이 있는 요소는 T_{Backup} 일 가능성이 높다. 그 이유는 블록체인과 융합을 하는 경우, 단순 저장장치 관련 프로세스가 아니라 블록체인 고유의 성능 비용이 소요되기 때문이다. 따라서, T_{Back} 를 $T_{BKStorageProc}$ 와 $T_{BKNetwork}$ 로 나누어 고려를 하는 것이 성능평가 모델에 적절하다고 볼 수 있다. 예측할 수 있는 것은, 데이터베이스의 성능보다는 비상 시 사용할 수 있는 보존성을 강화하는 하부 컴포넌트인 백업 시스템의 장치가 독립적인 블록체인망으로 대체되기 때문에 개발기간 및 개발노력에 영향을 미친다고 볼 수 있다. 다만, 이는 소요비용 증가로 예측되지만, 두 시스템 융합의 효과적인 결과는 무결성 그리고 지속성과 연관되어 있으므로 이는 정량적 분석을 실행하기가 어렵기때문에 정성적 정성 분석을 기반으로 논의를 하고자 한다[2, 13].

2.2 개념적 블록체인 성능분석 모델

블록체인을 데이터베이스 시스템과 융합하는 방법론 또는 새로운 형태의 구성에 대해서는 어느 특정 방법론이나 방향성이 있다고 할 수 없다. 이는 데이터베이스의 사용 범주 또는 아키텍처가 유사한 것에서부터 극적으로 상반되는 형태도 존재를 하며, 따라서 이를 일반화하는 것은 불가능하다고 할 수 있다. 이런 점을 고려할 때, 데이터베이스와 블록체인의 융합 사례에 대한 논의는 지속되어 왔다[15-17].

데이터베이스 시스템과 블록체인 융합시스템을 구성할 때, 블록체인 네트워크의 구조적 측면에서 새로운 데이터를 참여 블록체인 노드에 어떤 방법론을 통해 전파하는 지를 구별할 필요가 있다. 이는 실제적인 융합시스템의 성능에 영향을 미치기 때문이다. 기본적인 데이터 레코드 전파 방법은 트랜잭션 전파, 이벤트 발행과 구독, 채널 기반 전파, 앵커 피어 기반 전파(anchor peer based transmission) 등이 있다[18,19]. 블록체인 구조의 상이성을 극복하고 통일된 성능 분석 모델을 도출하기 위해 개념적으로 다음과 같은 RGBM(Resilient Block Generation Manager)을 정의한다. RGBM은 개념적으로 저장매체와 관련 관리 모듈을 규준화하여 성능분석을 단순화하기 위함이다. RGBM은 레질리언트 저장매체(resilient storage)를 사용하며 이를 위한 소프트웨어와 하드웨어를 포함하는 모듈로 구성한다[20]. 또한 RGBM은 여러 개의 단위 레코드를 컬렉션으로 집합화하고 압축하는 기능을 가진다고 가정한다.

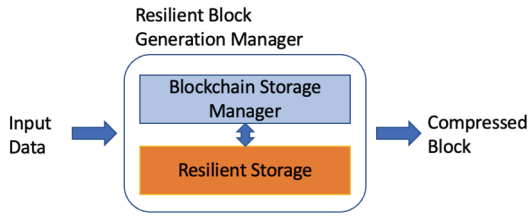


Fig. 2. Resilient Block Generation Manager.

이러한 RGBM 을 이용하면 제안하는 성능 분석 모델은 다양한 형태의 블록체인 네트워크를 다음과 같은 시나리오로 모델링할 수 있다. Fig. 3은 전체 블록체인을 하나의 성능요소로 구분할 수 있는 장점을 가진다.

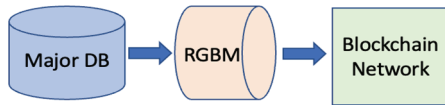


Fig. 3. Blockchain based Database System with a Gateway RGBM.

Fig. 4에서 보여주는 시나리오는 블록체인 각 노드 별로 독립적인 RGBM 을 가지고 있기 때문에 구성의 복잡성을 내포하는 모델을 제공할 수 있다.

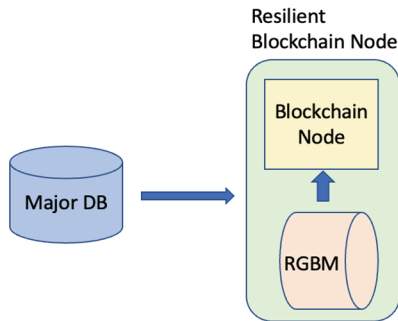


Fig. 4. Blockchain based Database System with independent RGBM's.

2.3 융합시스템의 정성적 성능분석

제한된 성능 분석 모델을 이용하여 위에서 주어진 식 2 는 다음과 같이 블록체인 적용으로 인한 새로운 성능요소 T_{BCCost} 를 포함하는 식으로 확장된다.

$$T_{Backup_BC} = T_{BKStoreProc} + T_{BKNetwork} + T_{BCCost} \quad (3)$$

식 3을 통하여 블록체인과 데이터베이스 시스템을 융

합한 시스템은 수행시간 측면에서 T_{BCCost} 를 위에서 제안한 두 가지 시나리오 중에서 블록체인 레코드 전파 방법론을 고려하여 선택하여 성능 분석 모델을 구성할 수 있다. 이를 통해 도출된 성능 분석 모델은 단순화된 분석을 통하여 수행 시간 측면의 예측을 가능하게 할 수 있다.

성능 분석 모델에서 전술한 수행 시간 측면 외에 우리가 고려하여야 하는 것은 저장 매체의 확장으로 인한 비용 분석일 것이다. 기본적으로 블록체인과의 융합을 제외한 기본 데이터베이스 시스템의 로그 저장 매체 비용을 단위 1로 가정하면, 레질리언트 저장 시스템은 작은 추가 비용을 부가하지만, 블록체인 네트워크에 포함된 노드의 범위성(*scalability*)을 n 이라고 하면 저장매체의 비용은 $O(n)$ 복잡성을 가질 수 밖에 없다. 또한 이러한 추가적인 비용과 기대할 수 있는 성능은 정량적 규정이 어려우므로, 단순화 시켜서 n 배의 추가적인 비용과 선행적인 관계의 무결성과 지속성 속성의 개선을 기대할 수 있다. 이를 토대로 융합시스템은 블록체인 노드들의 네트워크 멤버 구성 정책에 의해 전체적인 비용이 증가할 수 밖에 없다는 것을 알 수 있다.

3. 결 론

본 논문에서는 데이터베이스 시스템과 블록체인 네트워크의 융합을 통해 도출할 수 있는 성능 개선 분석을 수행 시간과 저장 매체 측면에서 개념적인 모델을 통하여 논의 하였다. 특히 데이터베이스 시스템의 로그 컴포넌트를 블록체인 네트워크와 연계하는 개념적 구조를 통하여 비용과 예상되는 개선된 시스템 성능을 분석하였다. 분석을 통해 예측할 수 있는 것은 융합 시스템의 구성에 따라 서 시스템의 트랜잭션 처리 시간은 최소화할 수 있는 가능성이 있음을 알 수 있었다. 하지만, 블록체인의 기본적인 구조에서 피할 수 없는 추가적인 블록체인 네트워크의 저장 시스템의 확대와 이에 따른 비용 증가 그리고 예측할 수 있는 데이터베이스 시스템의 무결성과 지속성 개선은 정량성을 확증할 수 없다는 점을 확인할 수 있었다.

감사의 글

이 연구는 2023학년도 단국대학교 대학연구비 지원으로 연구되었음.

참고문헌

1. Ullman, J. D., and Widom, J., A First Course in Database Systems. Pearson Education, 2008.

2. Osman, R., and Knottenbelt, W., "Database System Performance Evaluation Models: A Survey," *Performance Evaluation*, Vol. 69, No. 10, pp. 471-493, 2012.
3. Nakamoto, S., *Bitcoin: A Peer-to-Peer Electronic Cash System*, <https://bitcoin.org/bitcoin.pdf>, 2008. Accessed: 2023, 5, 31.
4. Gamage, H. T. M. et al., "A Survey on Blockchain Technology Concepts, Applications, and Issues," *SN Computer Science* 1, Vol. 114, 2020.
5. Dai, H., "Blockchain Challenges and Opportunities: A Survey," *International J. of Web and Grid Services*, Vol. 14, No.4, pp. 352-375, 2018.
6. An, J., Na, S., and Park, Y., "Smart Contract's Hierarchical Rules Modularization and Security Mechanism," *KSDT*, Vol. 18, No.1, pp. 74-78, 2019.
7. Zheng, Z., Xie, S., Dai, H., Chen, X., and Wang, H., "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," in *IEEE International Congress on Big Data*, pp. 557-564, 2017.
8. Lavanya, B.M., "Blockchain Technology Beyond Bitcoin: An Overview," *International J. of Computer Science and Mobile Applications*, Vol. 6, No. 1, pp. 76-80, 2018.
9. Shah, J., and Parveen, S., "Understanding the Blockchain Technology Beyond Bitcoin," *LNME*, pp. 499-516, 2021.
10. Sharma, N. et al., "Introduction to blockchain and distributed systems—fundamental theories and concepts," in *Blockchain for Smart Cities*, Elsevier, pp. 183-210, 2021.
11. Elmasri, R.A., and Shamkant B.N., *Fundamentals of Database Systems*, 5th Edition, 2006.
12. Petrov, A., *Database Internals: A Deep Dive into How Distributed Data Systems Work*, O'Reilly Media, 2019.
13. Lee, S. et al., "X-SSD: A Storage System with Native Support for Database Logging and Replication," in *Proceedings of the 2022 International Conference on Management of Data*, 2022.
14. Nicola, M., and Matthias J., "Performance Modeling of Distributed and Replicated Databases," *IEEE Trans. Knowl. Data Eng.*, Vol. 12, pp. 645-672, 2000.
15. O'Leary, D., "Configuring Blockchain Architectures for Transaction Information in Blockchain Consortiums: The Case of Accounting and Supply Chain Systems," *Intelligent Systems in Accounting, Finance and Management*, Vol. 24, No. , pp. 138-147, 2017.
16. Chowdhury, M.J.M., et al., "Blockchain Versus Database: A Critical Analysis," in *17th IEEE International Conf. On Trust, Security and Privacy In Computing and Communications* 12th IEEE International Conference On Big Data Science And Engineering, pp. 1348-1353, 2018.
17. Raikwar, M., et al., "Trends in Development of Databases and Blockchain," in *Seventh International Conference on Software Defined Systems*, pp. 177-182, 2020.
18. Gilani, K., Bertin, E., Hatin, J. and Crespi, N., "A Survey on Blockchain-based Identity Management and Decentralized Privacy for Personal Data," in *2nd Conference on Blockchain Research & Applications for Innovative Networks and Services*, pp. 97-101, 2020.
19. Pandey, R., et al., *Distributed computing to blockchain: Architecture, technology, and applications*, Academic Press, 2023.
20. Lee, Y., et al., "Hydra: Resilient and Highly Available Remote Memory," in *USENIX FAST*, 2022.

접수일: 2023년 6월 1일, 심사일: 2023년 6월 14일,
 게재확정일: 2023년 6월 21일