(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0018876
(43) 공개일자 2023년02월07일

(51) 국제특허분류(Int. Cl.)
G06F 16/182 (2019.01) *G06F 16/11* (2019.01)
G06F 16/13 (2019.01) *G06F 3/06* (2006.01)
H04L 9/06 (2006.01)
(52) CPC특허분류
G06F 16/1834 (2019.01)
G06F 16/122 (2019.01)
(21) 출원번호 10-2021-0100892
(22) 출원일자 2021년07월30일
심사청구일자 없음

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
박수용
서울특별시 마포구 독막로 18 길 5, 101동 1001호 (상수동, 상수두산위브아파트)
이열국
서울특별시 동대문구 한천로63길 10 이문e편한세상아파트 101동 2301호
(74) 대리인
이지연

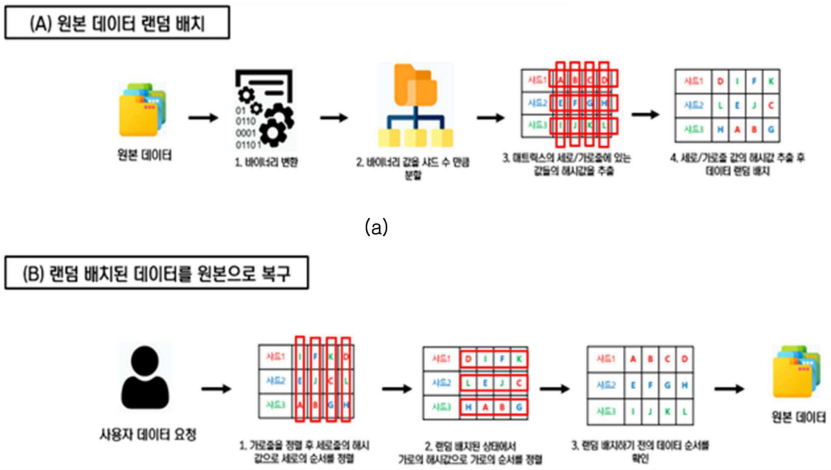
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **샤드를 기반으로 한 데이터 분산 저장 방법 및 이를 이용한 블록체인 네트워크 시스템**

(57) 요 약

본 발명은 블록체인 네트워크 시스템에서의 샤드를 기반으로 한 데이터 분산 저장 방법에 관한 것이다. 상기 데이터 분산 저장 방법은, (a) 원본 데이터를 인코딩하고, 인코딩된 데이터를 매트릭스로 변환시키고, 매트릭스의 가로줄에 대한 제1 해시값을 구한 후, 가로줄의 데이터를 랜덤으로 재배치하고, 매트릭스의 세로줄에 대한 제2 해시값을 구한 후, 세로줄의 데이터를 랜덤으로 재배치하고, 매트릭스의 랜덤 재배치된 데이터를 샤드들에 대응되게 분할하고, 분할된 데이터들과 제1 및 제2 해시값들을 샤드들에 분산 저장하는 단계 및 (b) 사용자의 요청에 따라, 샤드들로부터 랜덤 재배치된 데이터 및 해시값들을 제공받고, 랜덤 재배치된 데이터들과 해시값들을 이용하여 원본 데이터를 복원하는 단계;를 구비한다.

대 표 도 - 도2



(52) CPC특허분류
G06F 16/137 (2019.01)
G06F 3/064 (2013.01)
H04L 9/0643 (2013.01)
H04L 9/50 (2022.05)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103334
과제번호	2017-0-01628-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	대학 ICT 연구지원센터지원사업
연구과제명	적응형 블록체인 플랫폼 기술 개발 및 전문 인력 양성
기 여 율	1/1
과제수행기관명	서강대학교산학협력단
연구기간	2017.06.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

복수 개의 샤드들을 구비하는 블록체인 네트워크 시스템에서의 데이터 분산 저장 방법에 있어서,

(a) 원본 데이터를 인코딩하고 매트릭스로 변환시킨 후, 매트릭스의 각 데이터들에 대한 해시값을 구하고 랜덤 재배치하고, 랜덤 재배치된 데이터들을 사전 설정된 샤드들에 대응되게 분할시킨 후 각 샤드들에 분산 저장하는 단계; 및

(b) 사용자의 요청에 따라, 샤드들로부터 랜덤 재배치된 데이터 및 해시값들을 제공받고, 랜덤 재배치된 데이터들과 해시값들을 이용하여 원본 데이터를 복원하는 단계;

를 구비하는 것을 특징으로 하는 샤드를 기반으로 한 데이터 분산 저장 방법.

청구항 2

제1항에 있어서, 상기 (a) 단계는,

(a1) 원본 데이터를 인코딩하는 단계;

(a2) 인코딩된 데이터를 매트릭스의 형태로 변환시키되, 상기 매트릭스의 가로 및 세로 방향 중 어느 한 방향은 사전 설정된 샤드의 개수에 대응되도록 구성된 것을 특징으로 하는 단계;

(a3) 매트릭스의 제1 방향을 따라 각 줄에 대한 제1 해시값을 구한 후, 제1 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 단계;

(a4) 매트릭스의 제2 방향을 따라 각 줄에 대한 제2 해시값을 구한 후, 제2 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 단계; 및

(a5) 매트릭스의 랜덤 재배치된 데이터를 샤드들에 대응되게 분할하고, 분할된 데이터들과 제1 및 제2 해시값들을 샤드들에 분배하여 저장하는 단계;

를 구비하는 것을 특징으로 하는 샤드를 기반으로 한 데이터 분산 저장 방법.

청구항 3

제2항에 있어서, 상기 (a2) 단계는,

인코딩된 데이터를 매트릭스의 형태로 변환시키고, 매트릭스를 사전 설정된 크기의 블록으로 분할하는 것을 특징으로 하며,

상기 (a3), (a4), (a5) 단계는 매트릭스의 각 블록에 대하여 수행하는 것을 특징으로 하는 샤드를 기반으로 한 데이터 분산 저장 방법.

청구항 4

제1항에 있어서, 상기 (b) 단계는,

(b1) 사용자의 요청에 따라, 각 샤드들로부터 랜덤 재배치된 데이터 및 제1 해시값과 제2 해시값을 제공받는 단계;

(b2) 상기 랜덤 재배치된 데이터들을 매트릭스의 형태로 나열하는 단계;

(b3) 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제2 해시값을 비교하여, 제2 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 단계;

(b4) 제2 방향을 따라 재정렬된 매트릭스에 대하여, 매트릭스의 제1 방향을 따라 각 줄의 데이터들을 재배치하

고, 재배치된 데이터들의 해시값과 상기 제1 해시값을 비교하여, 제1 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 단계; 및

(b5) 상기 재정렬된 매트릭스의 데이터들을 디코딩하여 원본 데이터를 획득하는 단계;

를 구비하는 것을 특징으로 하는 샤드를 기반으로 한 데이터 분산 저장 방법.

청구항 5

제4항에 있어서, 상기 (b1) 단계는 각 샤드들로부터 복수 개의 블록에 대한 랜덤 재배치된 데이터 및 제1 해시값과 제2 해시값을 제공받는 것을 특징으로 하며,

상기 (b2), (b3), (b4) 단계는 각 블록들에 대하여 반복 수행되는 것을 특징으로 하며,

상기 (b5) 단계는 재정렬된 블록들을 순서대로 나열하여 매트릭스를 완성하고, 상기 매트릭스의 데이터들을 디코딩하여 원본 데이터를 획득하는 것을 특징으로 하는 샤드를 기반으로 한 데이터 분산 저장 방법.

청구항 6

제1항에 있어서, 상기 샤드는

블록체인 네트워크의 노드 또는 블록체인 네트워크의 노드 그룹으로 이루어진 것을 특징으로 하는 샤드를 기반으로 한 데이터 분산 저장 방법.

청구항 7

복수 개의 노드들을 구비하는 블록체인 네트워크 시스템에 있어서,

노드 또는 노드 그룹으로 이루어지는 복수 개의 샤드들이 사전 설정되고,

각 노드는

원본 데이터를 인코딩하고 매트릭스로 변환시킨 후, 매트릭스의 각 데이터들에 대한 해시값을 구하고 랜덤 재배치하고, 랜덤 재배치된 데이터들을 사전 설정된 샤드들에 대응되게 분할시킨 후 각 샤드들에 분산 저장하는 데이터 저장부; 및

사용자의 요청에 따라, 샤드들로부터 랜덤 재배치된 데이터 및 해시값들을 제공받고, 랜덤 재배치된 데이터들과 해시값들을 이용하여 원본 데이터를 복원하는 데이터 복원부;를 구비하여,

데이터를 샤드들에 분산 저장하는 것을 특징으로 하는 블록체인 네트워크 시스템.

청구항 8

제7항에 있어서, 상기 데이터 저장부는,

원본 데이터를 인코딩하는 인코딩 모듈;

인코딩된 데이터를 매트릭스의 형태로 변환시키되, 상기 매트릭스의 가로 및 세로 방향 중 어느 한 방향은 사전 설정된 샤드의 개수에 대응되도록 구성된 것을 특징으로 하는 매트릭스 변환 모듈;

매트릭스의 제1 방향을 따라 각 줄에 대한 제1 해시값을 구한 후, 제1 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 제1 랜덤 배치 모듈;

매트릭스의 제2 방향을 따라 각 줄에 대한 제2 해시값을 구한 후, 제2 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 제2 랜덤 배치 모듈; 및

매트릭스의 랜덤 재배치된 데이터를 샤드들에 대응되게 분할하고, 분할된 데이터들과 제1 및 제2 해시값들을 샤드들에 분배하여 저장하는 분산 저장 모듈;

을 구비하는 것을 특징으로 하는 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템.

청구항 9

제8항에 있어서, 상기 매트릭스 변환 모듈은,

인코딩된 데이터를 매트릭스의 형태로 변환시키고, 매트릭스를 사전 설정된 크기의 블록으로 분할하는 것을 특징으로 하며,

상기 제1 및 제2 랜덤 배치 모듈 및 분산 저장 모듈은 매트릭스의 각 블록에 대하여 수행하는 것을 특징으로 하는 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템.

청구항 10

제7항에 있어서, 상기 데이터 복원부는,

사용자의 요청에 따라, 각 샤드들로부터 랜덤 재배치된 데이터 및 제1 해시값과 제2 해시값을 제공받는 데이터 수집 모듈;

상기 랜덤 재배치된 데이터들을 매트릭스의 형태로 나열하는 매트릭스 정렬 모듈;

매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제2 해시값을 비교하여, 제2 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 제1 데이터 정렬 모듈;

제2 방향을 따라 재정렬된 매트릭스에 대하여, 매트릭스의 제1 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제1 해시값을 비교하여, 제1 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 제2 데이터 정렬 모듈; 및

상기 재정렬된 매트릭스의 데이터들을 디코딩하여 원본 데이터를 획득하는 데이터 복원 모듈;

을 구비하는 것을 특징으로 하는 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 샤드를 기반으로 한 데이터 분산 저장 방법 및 이를 이용한 블록체인 네트워크 시스템에 관한 것으로서, 더욱 구체적으로는 원본 데이터를 인코딩한 후 랜덤 배치하여 각 샤드에 분할 저장하고, 사용자의 요청에 따라 복원 및 검증할 수 있도록 한 데이터 분산 저장 방법 및 이를 이용한 블록체인 네트워크 시스템에 관한 것이다.

배 경 기 술

[0002] 블록체인은 데이터의 신뢰성과 효율성을 높일 수 있는 4차 산업혁명의 대표적인 기술 중 하나이다. 이러한 블록체인 기술에 있어서, 블록체인 프라이버시 보호를 위한 연구가 활발하게 진행되고 있으며, 그룹 서명(Group-signature), 링 서명(Ring-signature), 블라인드 서명(Blind-signature), zk-SNARKs 를 이용한 방식들이 다양하게 연구되고 있다. 전술한 기존의 방법들은 검증자가 별도로 존재하기 때문에, 검증에 대한 의존도와 특정 참여자에게 권한이 집중되는 중앙화 정도가 높은 문제점이 있다.

[0003] 한편, 샤딩(Sharding)은 데이터베이스에서 사용되는 개념으로써 대용량 데이터를 처리하기 위해 테이블을 수평 분할하여 데이터를 분산 저장하고 처리하기 위하여 사용되어 왔다. 샤드는 데이터의 수평 분할을 뜻하며, 분할된 데이터의 각각의 묶음을 의미한다. 그리고, 블록체인 분야에 있어서, 일반적으로 샤딩은 전체 네트워크를 나눈 뒤 트랜잭션을 영역별로 저장하고 이를 병렬적으로 처리하여 블록체인에 확장성을 부여하는 기술로 정의된다. 단일 데이터베이스에 저장할 수 없는 큰 데이터를 구간별로 쪼개어 나눔으로써 노드의 데이터를 검증한다. 샤딩은 필요한 데이터만 빠르게 조회가 가능하기 때문에, 쿼리가 가볍지만 샤드 내의 전송이 아닌 여러 샤드간의 전송은 절차가 복잡하고 느리다. 여러 개의 샤드로 나눌수록 샤드간 데이터를 어떻게 참조하고 검증할지에 대한 문제가 발생하여 알고리즘이 복잡해진다. 이러한 샤딩으로 나눠진 블록의 구간을 샤드(Shard)라 한다. 도 1은 샤딩으로 나누어진 샤드를 도시한 모식도이다.

[0004] 샤딩에 있어서 가장 우선시되는 고려사항은 샤드의 생성이고, 특정 샤드에서 통제력을 가진 사용자로부터의 공

격 가능성을 막기 위해 어떤 노드들이 어떤 샤드안에 존재하는지를 한정하기 위한 메커니즘이 요구된다. 이론적으로 트랜잭션 처리량이 샤드의 개수에 따라 선형적으로 증가하고 확장성 문제 해결외에, 샤딩은 블록체인의 장점인 탈중앙화와 보안적인 속성을 지니고 있기 때문에 블록체인의 보안을 유지하는 것도 가능하다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국공개특허공보 제 10-2021-0072265호

(특허문헌 0002) 한국공개특허공보 제 10-2021-0003234호

(특허문헌 0003) 한국공개특허공보 제 10-2021-0003214호

발명의 내용

해결하려는 과제

[0006] 전술한 문제점을 해결하기 위하여, 본 발명의 목적은 검증자가 별도로 존재하지 않고 데이터 요청자가 직접 검증할 수 있어 검증 의존도가 낮고 중앙화 정도도 낮은 데이터 분산 저장 방법 및 이를 이용한 블록체인 네트워크 시스템을 제공하는 것이다.

과제의 해결 수단

[0007] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징에 따른 복수 개의 샤드들을 구비하는 블록체인 네트워크 시스템에서의 데이터 분산 저장 방법은, (a) 원본 데이터를 인코딩하고 매트릭스로 변환시킨 후, 매트릭스의 각 데이터들에 대한 해시값을 구하고 랜덤 재배치하고, 랜덤 재배치된 데이터들을 사전 설정된 샤드들에 대응되게 분할시킨 후 각 샤드들에 분산 저장하는 단계; 및 (b) 사용자의 요청에 따라, 샤드들로부터 랜덤 재배치된 데이터 및 해시값들을 제공받고, 랜덤 재배치된 데이터들과 해시값들을 이용하여 원본 데이터를 복원하는 단계;를 구비한다.

[0008] 전술한 제1 특징에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 상기 (a) 단계는, (a1) 원본 데이터를 인코딩하는 단계; (a2) 인코딩된 데이터를 매트릭스의 형태로 변환시키되, 상기 매트릭스의 가로 및 세로 방향 중 어느 한 방향은 사전 설정된 샤드의 개수에 대응되도록 구성된 것을 특징으로 하는 단계; (a3) 매트릭스의 제1 방향을 따라 각 줄에 대한 제1 해시값을 구한 후, 제1 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 단계; (a4) 매트릭스의 제2 방향을 따라 각 줄에 대한 제2 해시값을 구한 후, 제2 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 단계; 및 (a5) 매트릭스의 랜덤 재배치된 데이터를 샤드들에 대응되게 분할하고, 분할된 데이터들과 제1 및 제2 해시값들을 샤드들에 분배하여 저장하는 단계;를 구비하는 것이 바람직하다.

[0009] 전술한 제1 특징에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 상기 (a2) 단계는 인코딩된 데이터를 매트릭스의 형태로 변환시키고, 매트릭스를 사전 설정된 크기의 블록으로 분할하는 것을 특징으로 하며, 상기 (a3), (a4), (a5) 단계는 매트릭스의 각 블록에 대하여 수행하는 것이 더욱 바람직하다.

[0010] 전술한 제1 특징에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 상기 (b) 단계는 (b1) 사용자의 요청에 따라, 각 샤드들로부터 랜덤 재배치된 데이터 및 제1 해시값과 제2 해시값을 제공받는 단계; (b2) 상기 랜덤 재배치된 데이터들을 매트릭스의 형태로 나열하는 단계; (b3) 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제2 해시값을 비교하여, 제2 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 단계; (b4) 제2 방향을 따라 재정렬된 매트릭스에 대하여, 매트릭스의 제1 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제1 해시값을 비교하여, 제1 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 단계; 및 (b5) 상기 재정렬된 매트릭스의 데이터들을 디코딩하여 원본 데이터를 획득하는 단계;를 구비하는 것이 바람직하다.

[0011] 전술한 제1 특징에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 상기 (b1) 단계는 각 샤드들로부터 복수 개의 블록에 대한 랜덤 재배치된 데이터 및 제1 해시값과 제2 해시값을 제공받는 것을 특징으로 하며,

상기 (b2), (b3), (b4) 단계는 각 블록들에 대하여 반복 수행되는 것을 특징으로 하며, 상기 (b5) 단계는 재정렬된 블록들을 순서대로 나열하여 매트릭스를 완성하고, 상기 매트릭스의 데이터들을 디코딩하여 원본 데이터를 획득하는 것이 더욱 바람직하다.

[0012] 전술한 제1 특징에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 상기 샤드는 블록체인 네트워크의 노드 또는 블록체인 네트워크의 노드 그룹으로 이루어진 것이 바람직하다.

[0013] 본 발명의 제2 특징에 따른 복수 개의 노드들을 구비하는 블록체인 네트워크 시스템은 노드 또는 노드 그룹으로 이루어지는 복수 개의 샤드들이 사전 설정되고, 각 노드는, 원본 데이터를 인코딩하고 매트릭스로 변환시킨 후, 매트릭스의 각 데이터들에 대한 해시값을 구하고 랜덤 재배치하고, 랜덤 재배치된 데이터들을 사전 설정된 샤드들에 대응되게 분할시킨 후 각 샤드들에 분산 저장하는 데이터 저장부; 및 사용자의 요청에 따라, 샤드들로부터 랜덤 재배치된 데이터 및 해시값들을 제공받고, 랜덤 재배치된 데이터들과 해시값들을 이용하여 원본 데이터를 복원하는 데이터 복원부;를 구비하여, 데이터를 샤드들에 분산 저장하는 것을 특징으로 한다.

[0014] 전술한 제2 특징에 따른 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템에 있어서, 상기 데이터 저장부는, 원본 데이터를 인코딩하는 인코딩 모듈; 인코딩된 데이터를 매트릭스의 형태로 변환시키되, 상기 매트릭스의 가로 및 세로 방향 중 어느 한 방향은 사전 설정된 샤드의 개수에 대응되도록 구성된 것을 특징으로 하는 매트릭스 변환 모듈; 매트릭스의 제1 방향을 따라 각 줄에 대한 제1 해시값을 구한 후, 제1 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 제1 랜덤 배치 모듈; 매트릭스의 제2 방향을 따라 각 줄에 대한 제2 해시값을 구한 후, 제2 방향을 따라 각 줄의 데이터를 랜덤으로 재배치하는 제2 랜덤 배치 모듈; 및 매트릭스의 랜덤 재배치된 데이터를 샤드들에 대응되게 분할하고, 분할된 데이터들과 제1 및 제2 해시값들을 샤드들에 분배하여 저장하는 분산 저장 모듈;을 구비한다.

[0015] 전술한 제2 특징에 따른 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템에 있어서, 상기 매트릭스 변환 모듈은, 인코딩된 데이터를 매트릭스의 형태로 변환시키고, 매트릭스를 사전 설정된 크기의 블록으로 분할하는 것을 특징으로 하며, 상기 제1 및 제2 랜덤 배치 모듈 및 분산 저장 모듈은 매트릭스의 각 블록에 대하여 수행하는 것이 더욱 바람직하다.

[0016] 전술한 제2 특징에 따른 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템에 있어서, 상기 데이터 복원부는, 사용자의 요청에 따라, 각 샤드들로부터 랜덤 재배치된 데이터 및 제1 해시값과 제2 해시값을 제공받는 데이터 수집 모듈; 상기 랜덤 재배치된 데이터들을 매트릭스의 형태로 나열하는 매트릭스 정렬 모듈; 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제2 해시값을 비교하여, 제2 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 제1 데이터 정렬 모듈; 제2 방향을 따라 재정렬된 매트릭스에 대하여, 매트릭스의 제1 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제1 해시값을 비교하여, 제1 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬하는 제2 데이터 정렬 모듈; 및 상기 재정렬된 매트릭스의 데이터들을 디코딩하여 원본 데이터를 획득하는 데이터 복원 모듈;을 구비하는 것이 바람직하다.

발명의 효과

[0017] 본 발명에 따른 샤드를 기반으로 한 데이터 분산 저장 방법은, 검증자가 별도로 존재하지 않고 데이터 요청자가 직접 검증할 수 있기 때문에, 검증 의존도가 낮을 뿐만 아니라 중앙화 정도도 낮기 때문에 탈중앙화가 가능하다. 또한, 본 발명에 따른 샤드를 기반으로 한 데이터 분산 저장 방법은, 검증 의존도가 낮기 때문에 블록체인 네트워크에 악의적인 노드가 존재하더라도 원본 데이터의 복원이 불가능하다.

[0018] 본 발명에 따른 샤드를 기반으로 한 데이터 분산 저장 방법은, 각 샤드에 랜덤 배치된 원본 데이터가 분배되면 문자의 순서가 뒤섞여 있기 때문에 각 샤드에 속한 노드들은 블록체인에 저장된 데이터의 원본 데이터의 형태를 알 수 없다. 예를 들어, 각 샤드에 저장된 데이터가 유출된다고 해도 블록체인의 내부 노드들 뿐만 아니라 블록체인 네트워크 외부에서도 블록체인에 저장되어 있는 데이터들의 원본을 알 수 없다.

[0019] 원본 데이터를 알기 위해서는 모든 샤드가 보유하고 있는 해시값이 필요하며, 해당 해시값이 가로 해시값인지 세로 해시값인지도 알고 있어야 한다. 따라서, 만약 해시값이 악의적인 사용자에 의해 노출되더라도, 원본이 어떠한 규칙으로 배치되었는지를 알 수 없고, 규칙을 모르는 상태에서 원본 데이터를 찾아내기 위한 조합의 경우의 수가 산술적으로 굉장히 큰 수이므로, 원본 데이터를 찾아내는 것은 실제로 거의 불가능하다.

[0020] 또한, 만약 악의적인 사용자가 우연히 전체 데이터 블록 중 일부분을 복원시킨다 하더라도, 나머지 데이터에 대한 정보를 알 수 없기 때문에 전체 데이터의 원본을 완벽하게 알 수 없고 원본 데이터를 유추할 수 없기 때문에, 안전성이 매우 높다.

[0021] 또한, 본 발명에 따른 샤드를 기반으로 한 데이터 분산 저장 방법은, 원본 데이터를 요청한 사용자가 직접 데이터를 검증할 수 있기 때문에 기존의 방법들보다 중앙화 정도와 검증 의존도가 낮다.

[0022] 도 13은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 샤드 개수에 따른 최소 블록의 데이터 복원 시간을 나타낸 그래프이다. 도 13을 참조하면, 본 발명에 따른 샤드를 기반으로 한 데이터 분산 저장 방법은, 원본 데이터가 여러 개의 최소 블록으로 이루어져 있고 샤드의 개수가 늘어날수록 세로줄 랜덤 배치에 대한 경우의 수가 2~4팩토리얼까지 증가하게 되고, 그에 따른 세로줄의 원본값과 해시값을 찾기 위한 대조과정이 많아지기 때문에, 복원시간이 늘어나고 프라이버시가 높아지게 된다.

[0023] 도 14는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 샤드 개수와 원본 데이터의 크기에 따른 데이터 복원 시간을 나타낸 그래프이다. 도 14를 참조하면, 데이터 크기가 증가하면 데이터 복원 시간도 비례하여 증가함을 확인할 수 있으며, 샤드의 개수가 늘어날수록 복원 시간이 늘어나고, 샤드의 개수가 고정되어 있을 때 원본 데이터의 크기가 증가함에 따라 데이터의 복원 시간도 비례하여 증가함을 확인할 수 있다. 이는 원본 데이터가 수많은 최소블록으로 이루어져 있기 때문에, 샤드의 개수가 같을 때, 원본 데이터의 크기가 n배 증가하면 최소블록의 개수도 n배 증가하고 블록의 복원 시간도 비례하여 증가하기 때문이다.

[0024] 도 15는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 최소 블록의 크기에 따른 데이터 복원 시간을 나타낸 그래프이다. 도 15를 참조하면, 데이터 최소 블록의 크기가 증가할수록 원본의 해시값을 찾는 경우의 수와 복원 시간이 비례하여 늘어나지 않고 더 큰 폭으로 늘어나는 것을 확인할 수 있다.

[0025] 도 16은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 기존의 블라인드 서명과 zk-SNARKs와 본 발명에 따른 방법의 검증 시간을 비교한 그래프이다. 도 16을 참조하면, 기존의 방법들에 비하여 본 발명에 따른 방법의 검증 시간이 많이 개선되었음을 확인할 수 있다.

도면의 간단한 설명

[0026] 도 1은 샤딩으로 나누어진 샤드를 도시한 모식도이다.

도 2는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법을 도시한 흐름도이다.

도 3은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 인코딩 단계에서 사용되는 Base 64의 색인표이다.

도 4는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정의 매트릭스 가로줄의 해시값 추출 단계를 도시한 모식도이다.

도 5는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정의 매트릭스 세로줄의 해시값 추출 단계 및 랜덤 배치 단계를 도시한 모식도이다.

도 6은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정의 데이터 매트릭스 블록화 단계를 도시한 모식도이다.

도 7은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정의 원본 데이터 랜덤 배치 상태를 예시적으로 도시한 모식도이다.

도 8은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정에 대한 알고리즘을 나타내는 의사코드이다.

도 9는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 복원 과정에서 해시값을 이용한 원본 데이터를 찾는 단계를 도시한 모식도이다.

도 10은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 복원 과정에 대한 알고리즘을 나타내는 의사코드이다.

도 11은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 복원 과정을 도시한 모식도이다.

도 12는 본 발명의 바람직한 실시예에 따른 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템에 있어서, 각 노드의 구성을 도시한 블록도이다.

도 13은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 샤드 개수에 따른 최소 블록의 데이터 복원 시간을 나타낸 그래프이다.

도 14는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 샤드 개수와 원본 데이터의 크기에 따른 데이터 복원 시간을 나타낸 그래프이다.

도 15는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 최소 블록의 크기에 따른 데이터 복원 시간을 나타낸 그래프이다.

도 16은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 기존의 블라인드 서명과 zk-SNARKs와 본 발명에 따른 방법의 검증 시간을 비교한 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0027] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법 및 이를 구현한 블록체인 네트워크 시스템에 대하여 구체적으로 설명한다.

[0028] 본 발명에 따른 샤드를 기반으로 한 데이터 분산 저장 방법은 블록체인 노드 또는 네트워크 그룹을 샤드로 정의하고, 트랜잭션 또는 원본 데이터를 인코딩한 값을 분할하여 각 샤드에 저장하며, 사용자로부터 요청에 따라 각 샤드에 저장된 트랜잭션 및 데이터를 복원 및 검증하는 것을 특징으로 한다.

[0029] 도 2는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법을 도시한 흐름도이다.

[0030] 도 2를 참조하면, 본 발명에 따른 샤드 기반의 데이터 분산 저장 방법은, 도 2a의 데이터 랜덤 배치(random placement of data) 과정과 도 2b의 데이터 복원(restore to original data) 과정을 구비한다.

[0031] 상기 데이터 랜덤 배치 과정은, 원본 데이터를 유추 불가능하게 하여 프라이버시를 높이고 블록체인이 지닌 투명성으로 인하여 누구나 데이터를 확인할 수 없도록 하기 위한 것이다. 도 2a를 참조하면, 데이터 랜덤 배치 과정은, 먼저 원본 데이터를 Base32 인코딩 방식을 이용하여 인코딩한다. 다음, 블록체인 네트워크 시스템의 샤드 개수에 맞게 데이터를 분배하기 위하여, 데이터를 가로와 세로 N*M 형식으로 매트릭스화한다. 이때, 매트릭스의 세로줄은 샤드의 개수를 의미한다. 도 2의 예시는 샤드가 3개이기 때문에 세로줄이 3줄로 되어 있음을 알 수 있다. 이와 같이, 데이터 분배를 위한 매트릭스화가 완료되면, 매트릭스의 가로줄을 랜덤하게 재배치하고, 세로줄을 랜덤하게 재배치하며, 매트릭스의 가로줄과 세로줄에 배치된 값들의 해시값을 추출함으로써, 원본 데이터 랜덤 배치 과정을 완료하게 된다. 이렇게 랜덤으로 배치가 완료된 데이터와 해시값들을 블록체인 네트워크의 각 샤드에 분배하여 저장한다.

[0032] 한편, 도 2b를 참조하면, 상기 복원 과정은, 사용자의 원본 데이터 요청에 따라, 원본 데이터로 복원 및 검증하는 과정으로서, 원본 데이터 랜덤 배치 과정의 역순으로 진행된다. 데이터 복원 과정은, 먼저 사용자로부터 블록체인에 저장된 데이터의 요청이 발생하면, 미리 추출했던 해시값들을 기반으로 랜덤으로 배치된 매트릭스를 원본의 형태로 재배치한다. 가로줄과 세로줄이 원본과 동일하게 재배치되었는지 여부를 해시값으로 확인한다.

[0033] 이하, 도 4 및 도 5를 참조하여 데이터 랜덤 배치 과정의 각 단계들에 대하여 보다 구체적으로 설명한다. 도 4는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정의 매트릭스 가로줄의 해시값 추출 단계를 도시한 모식도이며, 도 5는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정의 매트릭스 세로줄의 해시값 추출 단계 및 랜덤 배치 단계를 도시한 모식도이다.

[0034] 도 4를 참조하면, 데이터 랜덤 배치 과정은 먼저 원본 데이터 인코딩한다(Step 1). Step 1의 인코딩 단계는, 블록체인에 데이터를 저장하기 이전에, 문자, 이미지, 영상 데이터 등의 특정 데이터를 Base64 인코딩 방식을 활용하여 인코딩한다. Base64는 바이너리 데이터를 텍스트 데이터로 변경하는 인코딩 방식 중 하나로, 바이너리 데이터를 문자 코드에 영향받지 않는 64개의 아스키(ASCII) 영역의 문자들로 변경한다. 본 발명에 따른 방법에서 데이터를 인코딩하기 위하여 Base64 인코딩 방식을 사용하는 이유는 블록체인에서 가장 널리 사용되는 방식

이기 때문이다. Base64는 64진법을 의미하며, 알파벳 대소문자와 숫자, 그리고 +./를 사용한다. 도 3은 Base 64의 색인표이다.

[0035] 다음, 상기 인코딩된 데이터를 매트릭스화한다(Step 2). Step 2의 매트릭스화 단계는, 인코딩된 데이터를 소정 크기를 갖는 매트릭스의 형태로 변환한다. 데이터를 샤드들로 분배하기 위하여, 매트릭스의 가로 또는 세로의 크기는 샤드의 개수와 일치하여야 하며, 편의상 본 명세서에서는 매트릭스의 세로줄을 샤드의 개수와 동일하게 설정하여 설명한다. 예를 들면, 도 4의 Step1과 같이 특정 원본 데이터를 인코딩한 값이 'ABCDEFGHIJKL'이고, 블록체인 내 샤드의 개수가 3개라고 가정하면, 도 4의 Step2와 같이 인코딩한 데이터를 매트릭스화한다. 인코딩한 데이터를 샤드 1부터 샤드 3까지 분배한다.

[0036] 다음, 상기 매트릭스화된 데이터의 가로줄에 대한 해시값을 추출한다(Step 3). 도 4의 Step3과 같이. 매트릭스의 가로줄에 속한 값들에 SHA-1계열의 해시알고리즘을 사용하여 해시값을 추출하고, 매트릭스의 가로줄을 먼저 랜덤하게 배치한다(Step 4). 기존의 블록체인에서는 해시 알고리즘의 충돌 저항성이 높아 안정성이 좋은 SHA-256을 주로 사용하지만, 본 발명에서는 랜덤 배치로 인해 충돌 저항성이 만족되기 때문에 속도가 빠르고 용량 효율성이 높은 SHA-1 계열의 알고리즘을 사용한다.

[0037] 전술한 과정을 통해 매트릭스 안에서 가로줄을 랜덤 배치한 이후에, 매트릭스의 세로줄에 해당하는 해시값을 추출한다(Step 5). 세로줄에 대하여 각각 해시값이 생성되면, 도 5의 예시와 같은 4*3 매트릭스에서 총 7개의 해시값이 추출된다.

[0038] 이와 같이, 세로줄의 해시값을 추출한 후 세로줄을 다시 랜덤하게 배치함으로써(Step 6) 매트릭스의 랜덤 배치 과정은 종료한다. 본 발명에서 전술한 바와 같이, 가로줄과 세로줄에 대하여 랜덤 배치하는 것은 중요한 원본 데이터가 유출되거나 해킹되었을 때 데이터를 유추하지 못하도록 함으로써 프라이버시를 높이기 위함이다.

[0039] 도 4와 도 5에 예시된 데이터 매트릭스는 크기가 작기 때문에 복원 시간에 대한 문제가 발생하지 않는다. 하지만, 샤드의 수가 고정된 상태에서 원본 데이터의 인코딩 값이 증가할수록 각 샤드에 할당되는 데이터 값이 많아지고 이에 따라 생성되는 매트릭스의 크기도 커지게 된다. 매트릭스 세로줄의 길이는 초기에 설정해놓은 블록체인의 샤드 개수를 의미하기 때문에 고정값을 가진다. 따라서, 데이터의 크기가 증가할수록 매트릭스의 가로줄의 길이가 길어지고 랜덤 배치된 데이터를 복원할 때 경우의 수가 기하급수적으로 증가하기 때문에, 복원 시간이 오래 걸리게 된다. 이러한 데이터 크기 증가에 따른 복원 시간 문제를 해결하기 위하여, 매트릭스의 블록화를 진행하는 것이 바람직하다.

[0040] 본 발명에 있어서, 매트릭스의 블록화란 데이터의 가로줄을 일정하게 잘라 여러 개의 블록으로 나누는 것을 의미한다. 만약 매트릭스의 블록화를 하지 않는 경우, 중복값이 없는 문자 데이터 33자리와 샤드 3개를 가정했을 때, $8.6*10^{37}$ 개 이상의 엄청나게 큰 경우의 수가 발생하게 되고 복원하는 데 많은 시간이 소요된다.

[0041] 도 6은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정의 데이터 매트릭스 블록화 단계를 도시한 모식도이다. 도 6a는 원본 데이터로서, 블록으로 나누지 않고 원본 데이터를 랜덤으로 배치한다면 엄청난 숫자의 경우의 수가 발생하게 된다. 도 6b는 블록화된 원본 데이터를 예시적으로 도시한 것으로서, 본 발명에서는 각 블록을 데이터를 이루는 '최소블록'으로 정의하며, 최소블록마다 해시값을 추출하게 된다.

[0042] 본 발명에 따른 방법은 데이터에 랜덤 배치를 적용함으로써, 블록체인에 저장된 원본 데이터를 유추할 수 없게 된다. 도 7은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 원본 데이터 랜덤 배치하지 않은 상태와 원본 데이터 랜덤 배치한 상태를 예시적으로 도시한 모식도이다. 도 7을 참조하면, '안녕하세요'의 다섯 글자를 인코딩하면, '7JWI64WV7ZWY7IS47JqU'라는 인코딩 값이 생성된다. 도 7(a)와 같이, 인코딩 값을 5개의 샤드에 랜덤 배치과정없이 분배하게 되면 각 글자를 유추할 수 있게 된다. 하지만, 도 7(b)와 같이, 인코딩값을 5개의 샤드에 랜덤 배치한 후 분배하면 원본 데이터에 대한 유추가 불가능하게 된다.

[0043] 도 8은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 랜덤 배치 과정에 대한 알고리즘을 나타내는 의사코드이다.

[0044] 이하, 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 복원 과정을 보다 구체적으로 설명한다.

[0045] 매트릭스 문자 재배치 기법을 통해 랜덤으로 섞인 원본 데이터를 복원하기 위하여 데이터 랜덤 배치 과정에서 추출해 두었던 가로줄과 세로줄의 해시값이 필요하다. 블록체인 접근 권한을 가진 사용자가 블록체인에 저장된

데이터를 요청하면 데이터 매트릭스 블록을 랜덤 배치했을 때의 역순으로 세로줄의 값이 먼저 정렬된다.

[0046] 도 9는 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 복원 과정에서 해시값을 이용한 원본 데이터를 찾는 단계를 도시한 모식도이다. 도 9를 참조하면, 원본 데이터의 가로줄 또는 세로줄의 값이 'ABCD'라고 가정했을 때 원본 'ABCD'의 해시값은 'fb2f85c88567f3c8ce9b799c7c54642d0c7b41f6' 이라 가정한다. 원본의 값이 도 6과 같이, 'BACD', 'DABC','CADB' 등의 값처럼 배치되어 있을 때 원본의 배치인 'ABCD'로 복원하기 위해서 다시 한번 랜덤 함수가 사용된다.

[0047] 'BACD','DABC','CADB' 등의 배치의 해시값을 생성하고 원본값이 저장된 가로줄 또는 세로줄의 해시값과 일일이 대조하여 일치되는지 여부를 판단한다. 이 과정은 블록체인의 합의 알고리즘인 작업증명과 유사한 방식이며, 정렬되어야 할 데이터의 길이가 길수록 해시값을 대조하는 시간이 많이 소요된다.

[0048] 도 10은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 복원 과정에 대한 알고리즘을 나타내는 의사코드이며, 도 11은 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 한 데이터 분산 저장 방법에 있어서, 데이터 복원 과정을 도시한 모식도이다.

[0049] 도 11을 참조하면, 사용자가 각 샤드에 데이터를 요청하면, 각 샤드들로부터 랜덤 배치된 데이터들과 해시값들을 제공받게 된다. 각 샤드로부터 제공된 데이터와 해시값을 이용하여, 세로줄의 해시값을 탐색하고(Step1) 세로줄의 해시값을 통해 랜덤 배치 이전으로 복구함으로써(Step2), 세로줄을 먼저 정렬한다. 세로줄이 정렬된 상태는 원본 데이터의 매트릭스에서 가로줄의 값만 랜덤으로 배치된 상태이기 때문에 가로줄의 해시값을 세로줄과 같은 방식으로 대조하는 과정(Step3 및 Step4)을 거쳐 데이터의 원본 값으로 복원이 완료된다(Step5). 이때 사용자는 각 샤드로부터 저장된 데이터와 해시값을 모두 요청하게 되며, 하나의 샤드에서라도 요청한 값을 받지 못하고 누락될 시 원본 데이터를 완전 복원하는 것이 불가능하게 된다. 데이터를 요청한 사용자가 중앙 기관없이 직접 데이터와 해시값을 토대로 원본 데이터를 검증할 수 있다.

[0050] 이하, 도 12를 참조하여 본 발명의 바람직한 실시예에 따른 샤드를 기반으로 하여 데이터 분산 저장하는 블록체인 네트워크 시스템에 대하여 설명한다. 본 발명에 따른 블록체인 네트워크 시스템은 노드 또는 노드 그룹으로 이루어진 샤드들을 사전 설정하여 구비한다.

[0051] 도 12는 본 발명의 바람직한 실시예에 따른 샤드 기반의 데이터 분산 저장하는 블록체인 네트워크 시스템에 있어서, 각 노드의 구성을 도시한 블록도이다. 도 12를 참조하면, 각 노드(10)는 데이터 저장부(11) 및 데이터 복원부(12)를 구비한다. 상기 데이터 저장부(11)는 원본 데이터를 인코딩하고 매트릭스로 변환시킨 후, 매트릭스의 각 데이터들에 대한 해시값을 구하고 랜덤 재배치하고, 랜덤 재배치된 데이터들을 사전 설정된 샤드들에 대응되게 분할시킨 후 각 샤드들에 분산 저장한다. 상기 데이터 복원부(12)는, 사용자의 요청에 따라, 샤드들로부터 랜덤 재배치된 데이터 및 해시값들을 제공받고, 랜덤 재배치된 데이터들과 해시값들을 이용하여 원본 데이터를 복원한다.

[0052] 상기 데이터 저장부(11)는, 원본 데이터를 인코딩하는 인코딩 모듈(110), 매트릭스 변환 모듈(112), 제1 랜덤 배치 모듈(114), 제2 랜덤 배치 모듈(116) 및 분산 저장 모듈(118)을 구비한다.

[0053] 상기 매트릭스 변환 모듈(112)은 상기 인코딩된 데이터를 매트릭스의 형태로 변환시키되, 상기 매트릭스의 가로 및 세로 방향 중 어느 한 방향은 사전 설정된 샤드의 개수에 대응되도록 구성된다. 상기 제1 랜덤 배치 모듈(114)은 매트릭스의 제1 방향을 따라 각 줄에 대한 제1 해시값을 구한 후, 제1 방향을 따라 각 줄의 데이터를 랜덤으로 재배치한다. 상기 제2 랜덤 배치 모듈(116)은 매트릭스의 제2 방향을 따라 각 줄에 대한 제2 해시값을 구한 후, 제2 방향을 따라 각 줄의 데이터를 랜덤으로 재배치한다.

[0054] 상기 분산 저장 모듈(118)은 매트릭스의 랜덤 재배치된 데이터를 샤드들에 대응되게 분할하고, 분할된 데이터들과 제1 및 제2 해시값들을 샤드들에 분배하여 저장한다.

[0055] 한편, 상기 매트릭스 변환 모듈은, 인코딩된 데이터를 매트릭스의 형태로 변환시키고, 매트릭스를 사전 설정된 크기의 블록으로 분할하는 것이 바람직하며, 이 경우 상기 제1 및 제2 랜덤 배치 모듈 및 분산 저장 모듈은 매트릭스의 각 블록에 대하여 수행한다.

[0056] 상기 데이터 복원부(12)는, 데이터 수집 모듈(120), 매트릭스 정렬 모듈(122), 제1 데이터 정렬 모듈(124), 제2 데이터 정렬 모듈(126), 데이터 복원 모듈(128)을 구비한다. 상기 데이터 수집 모듈(120)은 사용자의 요청에 따라, 각 샤드들로부터 랜덤 재배치된 데이터 및 제1 해시값과 제2 해시값을 제공받는다. 상기 매트릭스 정렬 모

듈(122)은 상기 랜덤 재배치된 데이터들을 매트릭스의 형태로 나열한다. 상기 제1 데이터 정렬 모듈(124)은 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제2 해시값을 비교하여, 제2 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬한다. 상기 제2 데이터 정렬 모듈(126)은 제2 방향을 따라 재정렬된 매트릭스에 대하여, 매트릭스의 제1 방향을 따라 각 줄의 데이터들을 재배치하고, 재배치된 데이터들의 해시값과 상기 제1 해시값을 비교하여, 제1 해시값과 일치하는 해시값을 갖는 데이터를 획득하고, 획득된 데이터들을 이용하여 매트릭스의 제2 방향을 따라 각 줄의 데이터들을 재정렬한다. 상기 데이터 복원 모듈(128)은 상기 재정렬된 매트릭스의 데이터들을 디코딩하여 원본 데이터를 획득한다.

[0057] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0058] 10 : 노드

11 : 데이터 저장부

110 : 인코딩 모듈

112 : 매트릭스 변환 모듈

114 : 제1 랜덤 배치 모듈

116 : 제2 랜덤 배치 모듈

118 : 분산 저장 모듈

12 : 데이터 복원부

120 : 데이터 수집 모듈

122 : 매트릭스 정렬 모듈

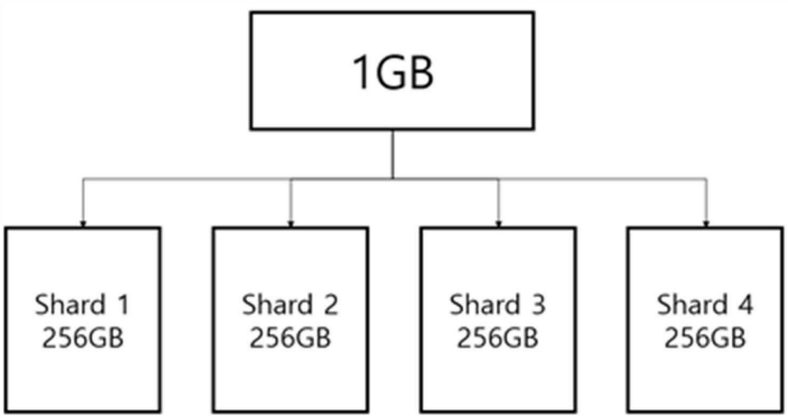
124 : 제1 데이터 정렬 모듈

126 : 제2 데이터 정렬 모듈

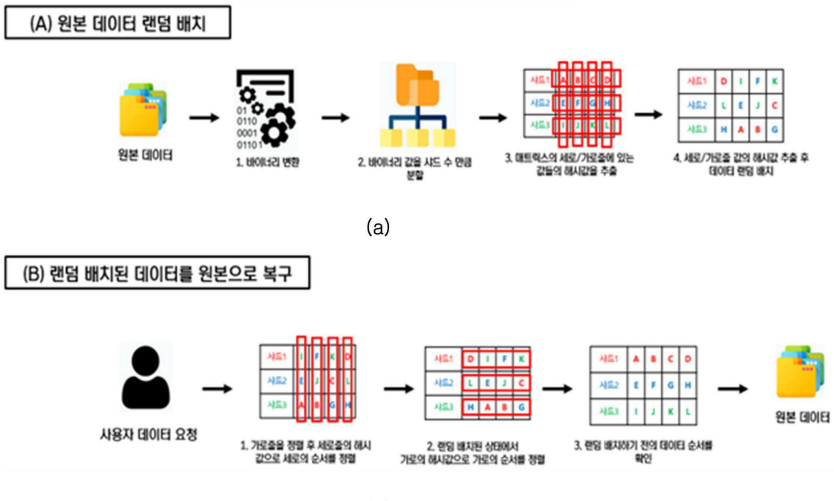
128 : 데이터 복원 모듈

도면

도면1



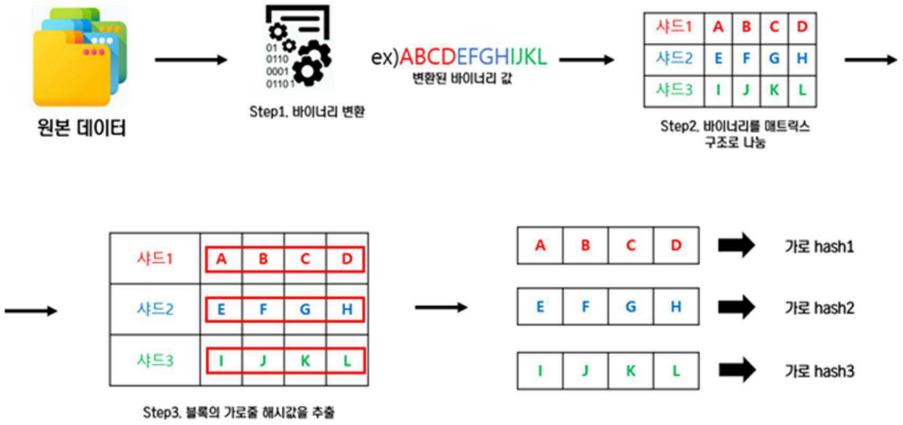
도면2



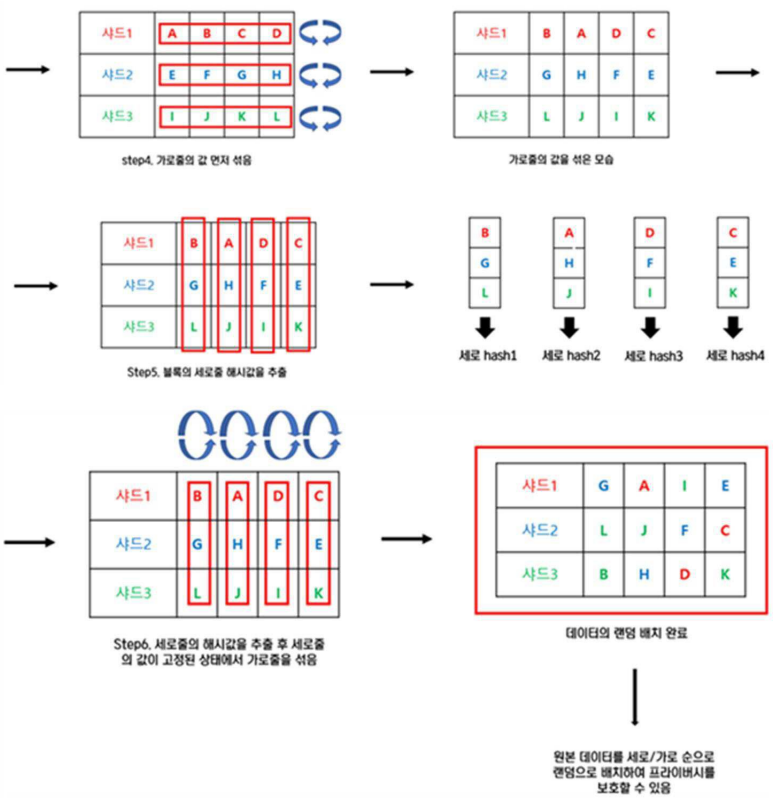
도면3

값	문자	값	문자	값	문자	값	문자
0	A	16	Q	32	g	48	w
1	B	17	R	33	h	49	x
2	C	18	S	34	i	50	y
3	D	19	T	35	j	51	z
4	E	20	U	36	k	52	0
5	F	21	V	37	l	53	1
6	G	22	W	38	m	54	2
7	H	23	X	39	n	55	3
8	I	24	Y	40	o	56	4
9	J	25	Z	41	p	57	5
10	K	26	a	42	q	58	6
11	L	27	b	43	r	59	7
12	M	28	c	44	s	60	8
13	N	29	d	45	t	61	9
14	O	30	e	46	u	62	+
15	P	31	f	47	v	63	/

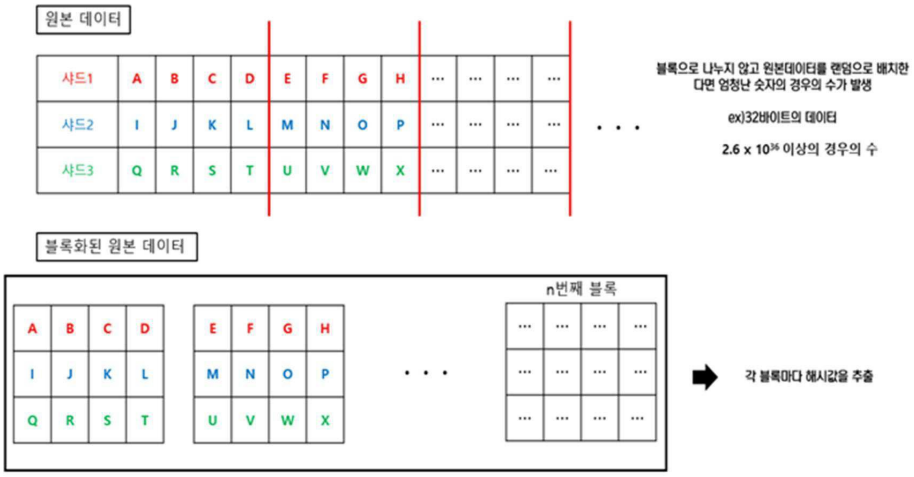
도면4


원본 데이터
Step1. 바이너리 변환
ex)ABCDEFGHIJKL
변환된 바이너리 값
샤드1 A B C D
샤드2 E F G H
샤드3 I J K L
Step2. 바이너리를 매트릭스 구조로 나눔
샤드1 A B C D
샤드2 E F G H
샤드3 I J K L
Step3. 블록의 가로줄 해시값을 추출
A B C D 가로 hash1
E F G H 가로 hash2
I J K L 가로 hash3

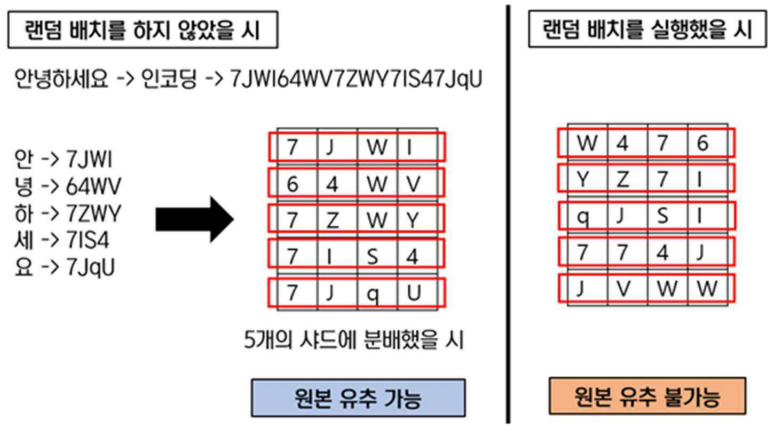
도면5


샤드1 A B C D
샤드2 E F G H
샤드3 I J K L
step4. 가로줄의 값 연지 섞음
샤드1 B A D C
샤드2 G H F E
샤드3 L J I K
가로줄의 값을 섞은 모습
Step5. 블록의 세로줄 해시값을 추출
B G L 세로 hash1
A H J 세로 hash2
D F I 세로 hash3
C E K 세로 hash4
Step6. 세로줄의 해시값을 추출 후 세로줄의 값이 고정된 상태에서 가로줄을 섞음
데이터의 랜덤 배치 완료
원본 데이터를 세로/가로 순으로 랜덤으로 배치하여 프라이버시를 보호할 수 있음

도면6



도면7



도면8

Algorithm 1 Random Placement of Data

Input: Original Data D, Number of shard N

Output: Shuffled D, RH, CH

1: $D \leftarrow \emptyset \in R^{N \times len(D)}$
2: **for** for each row from D **do**
3: $RH_i \leftarrow hash(D_i)$
4: shuffle D_i
5: **end for**
6: **for** for each columns from D **do**
7: $CH_i \leftarrow hash(D_j)$
8: shuffle D_j
9: **end for**
10: return D, RH, CH

도면9



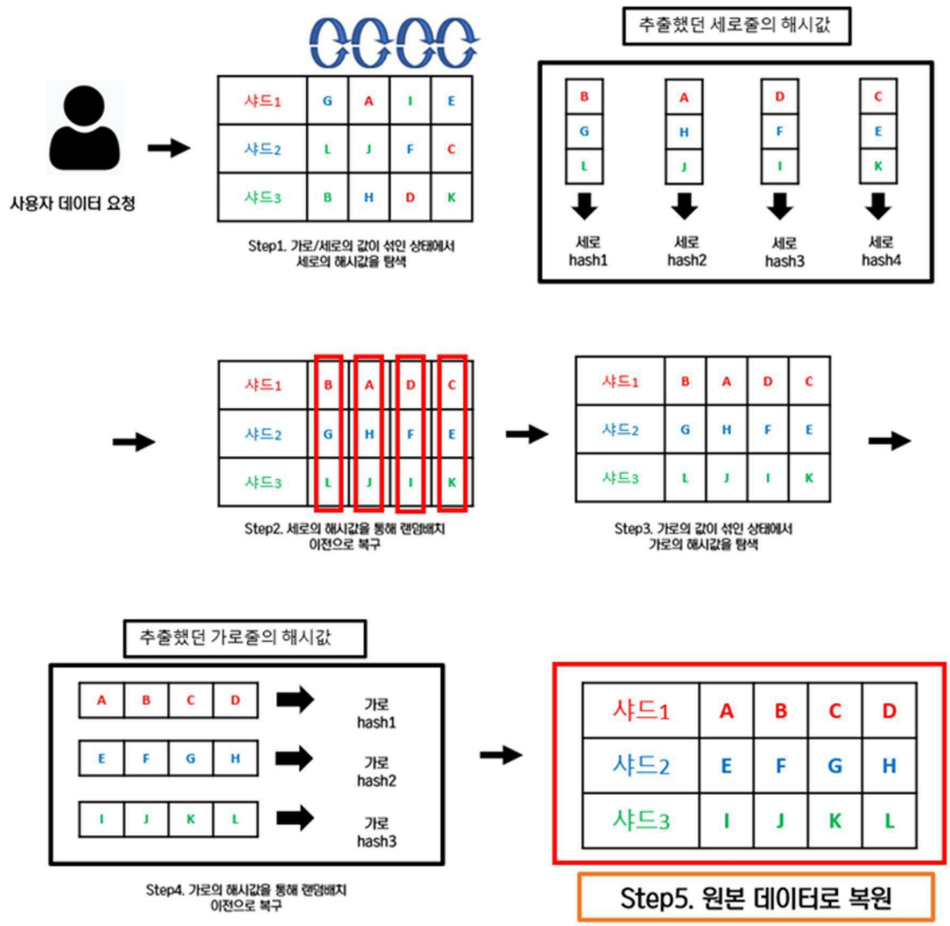
도면10

Algorithm 2 Restore to Original Data

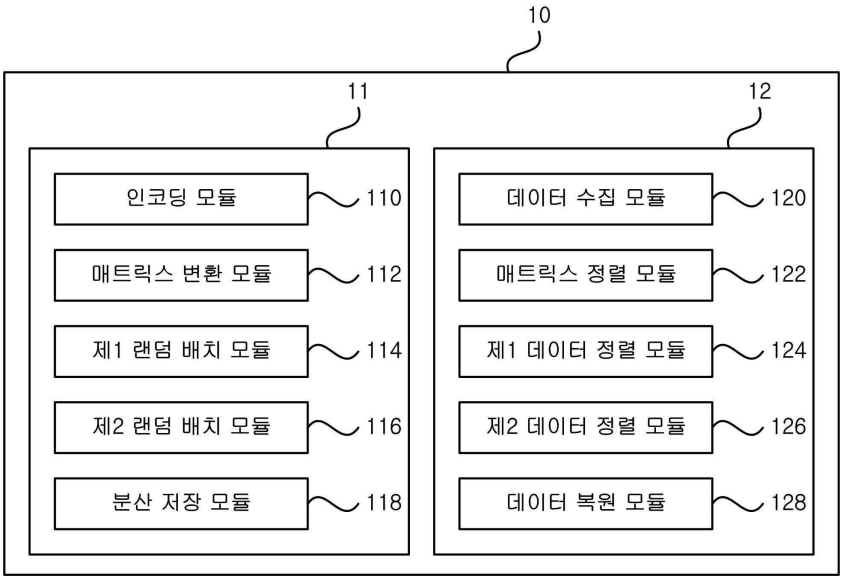
Input: $RH = \{RH_1, RH_2, ...RH_i\}$,
$CH = \{CH_1, CH_2, ...CH_k\}$
Output: Original Data D

set RD ← $RowData$
set ORH ← $OriginalRowHash$
for for each RD **do**
  shuffle RD_i
  if $hash(RD_i) == ORH_i$ **then**
    $D_i \leftarrow RD_i$
    break
  end if
end for
set CD ← $ColumnData$
set OCH ← $OriginalColumnHash$
for for each CD **do**
  shuffle CD_j
  if $hash(CD_j) == OCH_i$ **then**
    $D_j \leftarrow RD_j$
    break
  end if
end for
D ← concatenate(RD)
return D

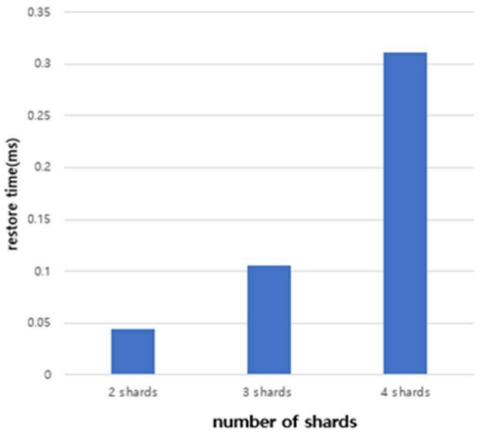
도면11


사용자 데이터 요청
샤드1 G A I E
샤드2 L J F C
샤드3 B H D K
Step1. 가로/세로의 값이 섞인 상태에서 세로의 해시값을 탐색
추출했던 세로줄의 해시값
B G L
A H J
D F I
C E K
세로 hash1
세로 hash2
세로 hash3
세로 hash4
샤드1 B A D C
샤드2 G H F E
샤드3 L J I K
Step2. 세로의 해시값을 통해 랜덤배치 이전으로 복구
샤드1 B A D C
샤드2 G H F E
샤드3 L J I K
Step3. 가로의 값이 섞인 상태에서 가로의 해시값을 탐색
추출했던 가로줄의 해시값
A B C D
E F G H
I J K L
가로 hash1
가로 hash2
가로 hash3
Step4. 가로의 해시값을 통해 랜덤배치 이전으로 복구
샤드1 A B C D
샤드2 E F G H
샤드3 I J K L
Step5. 원본 데이터로 복원

도면12

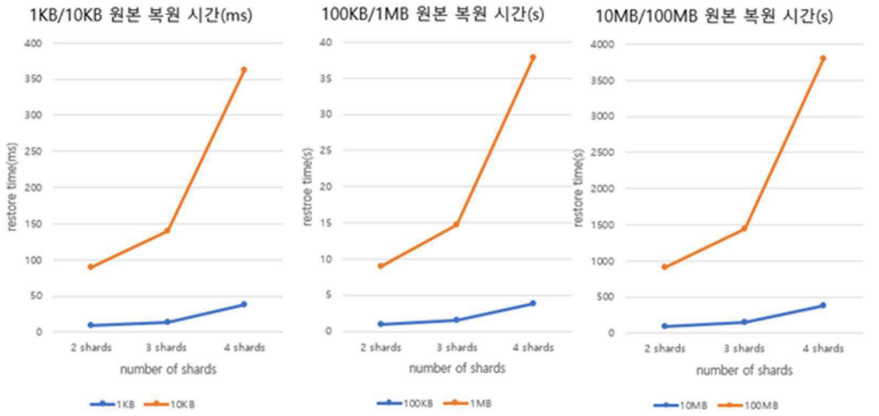

10
11
12
인코딩 모듈 110
매트릭스 변환 모듈 112
제1 랜덤 배치 모듈 114
제2 랜덤 배치 모듈 116
분산 저장 모듈 118
데이터 수집 모듈 120
매트릭스 정렬 모듈 122
제1 데이터 정렬 모듈 124
제2 데이터 정렬 모듈 126
데이터 복원 모듈 128

도면13



number of shards	2 shards	3 shards	4 shards
restore time(ms)	0.044	0.106	0.311

도면14

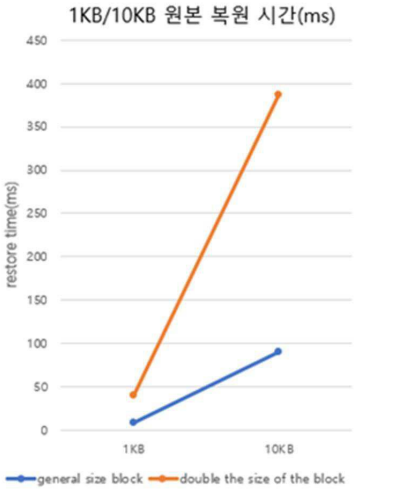


	2 shards	3 shards	4 shards
1KB	8.89ms	14.06ms	37.82ms
10KB	89.87ms	139.37ms	362.21ms

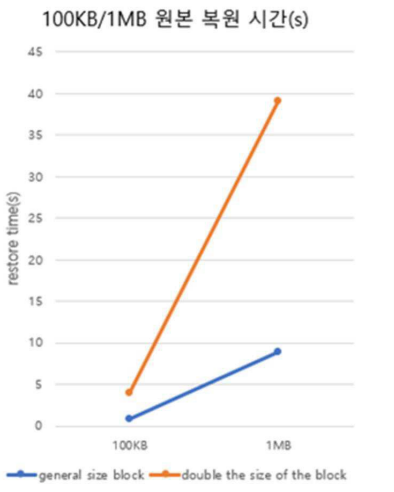
	2 shards	3 shards	4 shards
100KB	0.8865s	1.4895s	3.8521s
1MB	8.981s	14.661s	37.885s

	2 shards	3 shards	4 shards
10MB	89.764s	142.15s	385.14s
100MB	909.21s	1441.18s	3791.85s

도면15



	1KB	10KB
general size block	8.89ms	89.87ms
double the size of the block	39.26ms	386.48ms



	100KB	1MB
general size block	0.886s	8.98s
double the size of the block	3.968s	39.09s

도면16

