



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0103237
(43) 공개일자 2023년07월07일

(51) 국제특허분류(Int. Cl.)
G06F 9/50 (2018.01) H04L 67/1042 (2022.01)
(52) CPC특허분류
G06F 9/5066 (2013.01)
G06F 9/5077 (2013.01)
(21) 출원번호 10-2021-0193974
(22) 출원일자 2021년12월31일
심사청구일자 2021년12월31일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
정중문
서울특별시 용산구 이촌로87길 21, 106동 1502호 (이촌동, 이촌아파트)
서승우
서울특별시 서대문구 성산로 367-35, 1동 205호 (연희동, 샬롬하우스)
(74) 대리인
민영준

전체 청구항 수 : 총 20 항

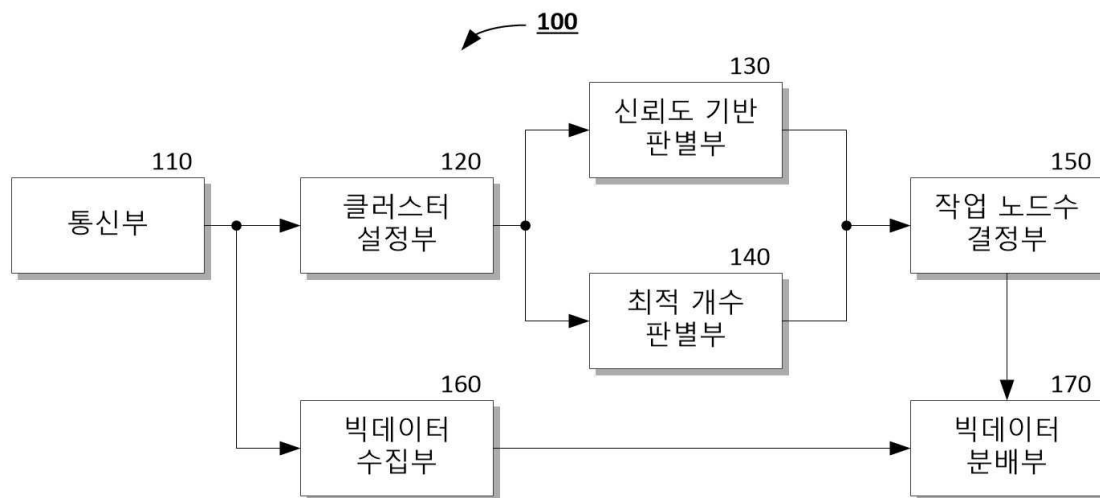
(54) 발명의 명칭 적응적 클러스터링을 이용한 빅데이터 처리 제어 장치 및 방법

(57) 요약

본 발명은 작업 명령에 따라 처리되어야 하는 빅데이터를 분산 처리하는 다수의 작업 노드와 빅데이터 처리 제어 장치를 포함하는 빅데이터 시스템에 있어서, 다수의 작업 노드에 대한 신뢰도를 계산하는 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하여 다수의 작업 노드를 계산된 최적 클러스터 개수의 클러스터

(뒷면에 계속)

대표도 - 도2



로 구분하고, 다수의 작업 노드 각각에 대해 획득된 신뢰도 정보를 기반으로 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 판별하며, 작업 명령에서 요구하는 시간인 목표 시간 이내에 빅데이터를 구분된 클러스터로 분산 처리할 수 있는 작업 노드의 개수인 최적 작업 노드 개수를 계산하고, 정직 작업 노드의 개수와 최적 작업 노드 개수에 따라 실제 빅데이터를 분산처리하는 작업 노드의 개수인 처리 노드 수를 결정하는 시스템 관리부 및 데이터를 인가받아 수집하여 빅데이터를 생성하고, 생성된 빅데이터를 처리 노드 수의 작업 노드가 처리하도록 분배하는 빅데이터 관리부를 포함하여, 보안성을 향상시키면서 작업 명령이 요구하는 목표 시간을 만족시킬 수 있도록 하며, 악의적 작업 노드를 빅데이터 처리에서 배제시킴으로써 작업 노드에 대한 보안 공격 상황에서도 빅데이터 처리 성능이 저하되는 것을 가능한 억제할 수 있는 빅데이터 처리 제어 장치 및 방법을 제공한다.

(52) CPC특허분류

H04L 67/1046 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711126081
과제번호	2018-0-01799-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	블록체인 비즈니스 서비스 기술 개발 및 인력양성
기 여 율	1/1
과제수행기관명	중앙대학교산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

작업 명령에 따라 처리되어야 하는 빅데이터를 분산 처리하는 다수의 작업 노드와 빅데이터 처리 제어 장치를 포함하는 빅데이터 시스템에 있어서,

상기 다수의 작업 노드에 대한 신뢰도를 계산하는 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하여 상기 다수의 작업 노드를 계산된 최적 클러스터 개수의 클러스터로 구분하고, 상기 다수의 작업 노드 각각에 대해 획득된 신뢰도 정보를 기반으로 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 판별하며, 상기 작업 명령에서 요구하는 시간인 목표 시간 이내에 빅데이터를 구분된 클러스터로 분산 처리할 수 있는 작업 노드의 개수인 최적 작업 노드 개수를 계산하고, 상기 정직 작업 노드의 개수와 상기 최적 작업 노드 개수에 따라 실제 빅데이터를 분산처리하는 작업 노드의 개수인 처리 노드 수를 결정하는 시스템 관리부; 및

데이터를 인가받아 수집하여 빅데이터를 생성하고, 생성된 빅데이터를 상기 처리 노드 수의 작업 노드가 처리하도록 분배하는 빅데이터 관리부를 포함하는 빅데이터 처리 제어 장치.

청구항 2

제1항에 있어서, 상기 시스템 관리부는

구분되는 클러스터의 개수에 따른 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하고, 계산된 최적 클러스터 개수에 따라 상기 다수의 작업 노드를 클러스터링하여 구분하며, 구분된 각 클러스터에 포함된 작업 노드 각각의 신뢰도에 따라 클러스터를 관리하는 클러스터 헤드 노드를 지정하는 클러스터 설정부;

구분된 각 클러스터의 클러스터 헤드 노드가 대응하는 클러스터 내의 작업 노드 각각에 대해 획득한 신뢰도 정보를 인가받아, 상기 다수의 작업 노드에서 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 확인하는 신뢰도 기반 판별부;

상기 빅데이터 처리 제어 장치가 상기 빅데이터를 각 클러스터로 분배하기 위해 소요되는 마스터 처리 시간과 각 클러스터가 분배된 빅데이터를 처리하기 위해 소요되는 클러스터 처리 시간을 계산하여 상기 작업 명령에 대응하여 빅데이터를 처리하는 빅데이터 처리 시간을 획득하고, 상기 빅데이터 처리 시간이 상기 목표 시간을 만족시킬 수 있도록 하는 최소의 작업 노드 개수인 최적 작업 노드 개수를 계산하는 최적 개수 판별부; 및

상기 정직 작업 노드의 개수와 상기 최적 작업 노드 개수를 비교하여 상기 처리 노드 수를 결정하는 작업 노드 수 결정부를 포함하는 빅데이터 처리 제어 장치.

청구항 3

제2항에 있어서, 상기 클러스터 설정부는

상기 빅데이터 시스템의 환경 변화를 분석하여 클러스터 재구성 여부를 판별하는 클러스터 재구성 판별부;

클러스터를 재구성해야 하는 것으로 판별되면, 전체 작업 노드의 개수에 따라 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하는 최적 클러스터 결정부;

계산된 최적 클러스터 개수에 따라 작업 노드를 다수의 작업 노드를 클러스터링하여 구분하는 작업 노드 클러스터링부; 및

각 클러스터에 포함된 작업 노드 각각의 이전 획득되어 저장된 신뢰도 정보를 확인하여, 가장 높은 신뢰도를 갖는 작업 노드를 해당 클러스터의 클러스터 헤드 노드로 선택하는 클러스터 헤드 선택부를 포함하는 빅데이터 처리 제어 장치.

청구항 4

제3항에 있어서, 상기 최적 클러스터 결정부는

상기 전체 작업 노드의 개수(n)에 따라 상기 최적 클러스터 개수(k_{opt})를 $nint(\sqrt{n})$ (여기서 $nint(x)$ 는 x에 가장 가까운 정수)으로 계산하는 빅데이터 처리 제어 장치.

청구항 5

제3항에 있어서, 상기 신뢰도 기반 판별부는

구분된 각 클러스터의 클러스터 헤드 노드가 대응하는 클러스터 내의 작업 노드 각각에 대해 획득한 신뢰도 정보를 인가받아 기지정된 방식으로 검증하는 신뢰도 정보 검증부;

검증된 신뢰도 정보에 따라 상기 다수의 작업 노드 각각을 상기 정직 작업 노드 또는 상기 악의적 작업 노드로 판별하는 악의적 노드 선별부; 및

판별된 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 확인하는 신뢰 노드수 확인부를 포함하는 빅데이터 처리 제어 장치.

청구항 6

제5항에 있어서, 상기 신뢰도 정보는

각 클러스터에서 상기 클러스터 헤드 노드가 다수의 작업 노드 각각이 동일 클러스터 내의 다른 작업 노드가 빅데이터를 처리한 결과를 분석하여 생성한 로컬 신뢰 의견을 수집하고, 수집된 로컬 신뢰 의견 사이의 유사도를 기반으로 주관 평판을 계산하여, 클러스터 내의 작업 노드 중 유사한 주관 평판을 갖는 과반 이상의 작업 노드를 그룹화하여 신뢰 정족수 집합을 획득하고, 획득된 상기 신뢰 정족수 집합 내의 작업 노드의 주관 평판을 기반으로 각 작업 노드가 다른 작업 노드의 빅데이터 처리한 행동을 평가한 행동 평판을 계산하며, 상기 로컬 신뢰 의견과 상기 행동 평판을 기반으로 각 작업 노드가 다른 작업 노드를 평가하는 행동에 대한 신뢰도 평판을 계산하여 획득하는 글로벌 평판인 빅데이터 처리 제어 장치.

청구항 7

제6항에 있어서, 상기 로컬 신뢰 의견(LTO)은 수학적식

$$LTO_{j,k} = \frac{p_{j,k}}{p_{j,k} + n_{j,k}}$$

(여기서 $LTO_{j,k}$ 는 클러스터 내의 제j 작업 노드(WN_j)가 제k 작업 노드(WN_k)에 대해 측정한 로컬 신뢰 의견, $p_{j,k}$ 는 기지정된 시간 동안 관측된 긍정적 행동 횟수로 정상적으로 빅데이터를 처리한 횟수이고, $n_{j,k}$ 는 부정적 행동 횟수로 빅데이터 처리에 요구되는 목표 시간(T_{object})을 초과하거나, 통신 오류가 발생하는 등의 비정상적인 행동이 발생한 횟수이다.)

상기 주관 평판(SR)은

$$SR_{w,u} = \sum_{j \in S_u} LTO_{j,u} \frac{HR_j \text{sim}(w,j)}{\sum_{j \in S_u} HR_j \text{sim}(w,j)}$$

(여기서 S_u 는 클러스터 내의 제u 작업 노드(WN_u)에 대한 로컬 신뢰 의견(LTO)을 전송한 작업 노드 집합을 나타내고, $\text{sim}(w,j)$ 는 제w 작업 노드(WN_w)와 제j 작업 노드(WN_j)가 보고한 로컬 신뢰 의견(LTO) 사이의 코사인 유사도를 나타내며, HR_j 는 로컬 신뢰 의견(LTO)을 생성한 각 작업 노드의 계층별 가중치로서 일반 작업 노드(WN)는 1이고, 클러스터 헤드 노드(CH)는 1보다 큰 기지정된 값을 갖는다.)

에 따라 계산되며,

상기 행동 평판(BR)은 수학적식

$$BR_{CH_i,u} = \frac{\sum_{w \in D} SR_{w,u}}{|D_{CH_i}|}$$

(여기서 $|D_{CH_i}|$ 는 제i 클러스터(CL_i)의 클러스터 헤드(CH_i)가 구성한 신뢰 정족수 집합(D)에 포함된 작업 노드의 개수를 나타낸다.)

에 따라 계산되고,

상기 신뢰도 평판(CR)은 수학적식

$$CR_{CH_i,u} = 1 - \sqrt{\frac{\sum_{j \in \{LTO_{u,j} \neq null\}} (LTO_{u,j} - BR_{CH_i,u})^2}{|j \in \{LTO_{u,j} \neq null\}|}}$$

에 따라 계산되는 빅데이터 처리 제어 장치.

청구항 8

제5항에 있어서, 상기 최적 개수 판별부는

상기 마스터 처리 시간을 상기 빅데이터 처리 제어 장치와 상기 클러스터 헤드 노드 사이의 변수 공유 시간과 데이터 전송 시간 및 데이터 전송 오류를 복구하기 위한 오류 복구 시간의 합으로 계산하는 전송 시간 계산부;

상기 클러스터 처리 시간을 상기 클러스터 헤드 노드와 동일 클러스터 내의 나머지 작업 노드 사이의 변수 공유 시간과 데이터 전송 시간, 데이터 전송 오류 및 작업 노드들의 빅데이터 처리 오류를 복구하기 위한 오류 복구 시간 및 데이터 처리 시간의 합으로 계산하는 작업 시간 계산부; 및

상기 마스터 처리 시간과 상기 클러스터 처리 시간의 합으로 계산되는 상기 빅데이터 처리 시간이 상기 목표 시간 이하가 되도록 하는 최소값을 최적 작업 노드 개수로서 획득하는 최적 개수 계산부를 포함하는 빅데이터 처리 제어 장치.

청구항 9

제8항에 있어서, 상기 작업 노드수 결정부는

상기 정적 작업 노드의 개수가 상기 최적 작업 노드 개수 이상이면, 상기 최적 작업 노드 개수를 상기 처리 노드 수로 결정하고, 상기 정적 작업 노드의 개수가 상기 최적 작업 노드 개수보다 작으면, 상기 정적 작업 노드의 개수를 상기 처리 노드 수로 결정하는 빅데이터 처리 제어 장치.

청구항 10

제1항에 있어서, 상기 빅데이터 관리부는

데이터를 인가받아 수집하여 빅데이터를 생성하고, 각 클러스터에서 처리된 빅데이터를 인가받아 수집하여 작업 명령을 인가한 사용자 단말로 전송하는 빅데이터 수집부; 및

빅데이터를 상기 처리 노드 수에 따라 분할하고, 분할된 빅데이터를 각 클러스터에 포함된 정적 작업 노드의 개수를 고려하여 클러스터 단위로 분배하는 빅데이터 분배부를 포함하는 빅데이터 처리 제어 장치.

청구항 11

빅데이터 시스템의 빅데이터 처리 제어 장치에서 수행되는 빅데이터 처리 제어 방법에 있어서,

작업 명령에 따라 처리되어야 하는 빅데이터를 분산 처리하는 다수의 작업 노드에 대한 신뢰도를 계산하는 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하여 상기 다수의 작업 노드를 계산된 최적 클러스터 개수의 클러스터로 구분하고, 상기 다수의 작업 노드 각각에 대해 획득된 신뢰도 정보를 기반으로 정적 작업 노드의 개수와 악의적 작업 노드의 개수를 판별하며, 상기 작업 명령에서 요구하는 시간인 목표 시간 이내에 빅데이터를 구분된 클러스터로 분산 처리할 수 있는 작업 노드의 개수인 최적 작업 노드 개수를 계산하고,

상기 정직 작업 노드의 개수와 상기 최적 작업 노드 개수에 따라 실제 빅데이터를 분산처리하는 작업 노드의 개수인 처리 노드 수를 결정하는 단계; 및

데이터를 인가받아 수집하여 빅데이터를 생성하고, 생성된 빅데이터를 상기 처리 노드 수의 작업 노드가 처리하도록 분배하는 단계를 포함하는 빅데이터 처리 제어 방법.

청구항 12

제11항에 있어서, 상기 처리 노드 수를 결정하는 단계는

구분되는 클러스터의 개수에 따른 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하고, 계산된 최적 클러스터 개수에 따라 상기 다수의 작업 노드를 클러스터링하여 구분하며, 구분된 각 클러스터에 포함된 작업 노드 각각의 신뢰도에 따라 클러스터를 관리하는 클러스터 헤드 노드를 지정하는 단계;

구분된 각 클러스터의 클러스터 헤드 노드가 대응하는 클러스터 내의 작업 노드 각각에 대해 획득한 신뢰도 정보를 인가받아, 상기 다수의 작업 노드에서 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 확인하는 단계;

상기 빅데이터 처리 제어 방법이 상기 빅데이터를 각 클러스터로 분배하기 위해 소요되는 마스터 처리 시간과 각 클러스터가 분배된 빅데이터를 처리하기 위해 소요되는 클러스터 처리 시간을 계산하여 상기 작업 명령에 대응하여 빅데이터를 처리하는 빅데이터 처리 시간을 획득하고, 상기 빅데이터 처리 시간이 상기 목표 시간을 만족시킬 수 있도록 하는 최소의 작업 노드 개수인 최적 작업 노드 개수를 계산하는 단계; 및

상기 정직 작업 노드의 개수와 상기 최적 작업 노드 개수를 비교하여 상기 처리 노드 수를 판별하는 단계를 포함하는 빅데이터 처리 제어 방법.

청구항 13

제12항에 있어서, 상기 클러스터 헤드 노드를 지정하는 단계는

상기 빅데이터 시스템의 환경 변화를 분석하여 클러스터 재구성 여부를 판별하는 단계;

클러스터를 재구성해야 하는 것으로 판별되면, 전체 작업 노드의 개수에 따라 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하는 단계;

계산된 최적 클러스터 개수에 따라 작업 노드를 다수의 작업 노드를 클러스터링하여 구분하는 단계; 및

각 클러스터에 포함된 작업 노드 각각의 이전 획득되어 저장된 신뢰도 정보를 확인하여, 가장 높은 신뢰도를 갖는 작업 노드를 해당 클러스터의 클러스터 헤드 노드로 선택하는 단계를 포함하는 빅데이터 처리 제어 방법.

청구항 14

제13항에 있어서, 상기 최적 클러스터 결정부는

상기 전체 작업 노드의 개수(n)에 따라 상기 최적 클러스터 개수(k_{opt})를 $nint(\sqrt{n})$ (여기서 $nint(x)$ 는 x 에 가장 가까운 정수)으로 계산하는 빅데이터 처리 제어 방법.

청구항 15

제13항에 있어서, 상기 악의적 작업 노드의 개수를 확인하는 단계는

구분된 각 클러스터의 클러스터 헤드 노드가 대응하는 클러스터 내의 작업 노드 각각에 대해 획득한 신뢰도 정보를 인가받아 기지정된 방식으로 검증하는 단계;

검증된 신뢰도 정보에 따라 상기 다수의 작업 노드 각각을 상기 정직 작업 노드 또는 상기 악의적 작업 노드로 판별하는 단계; 및

판별된 악의적 작업 노드의 개수와 정직 작업 노드의 개수를 확인하는 단계를 포함하는 빅데이터 처리 제어 방법.

청구항 16

제15항에 있어서, 상기 신뢰도 정보는

각 클러스터에서 상기 클러스터 헤드 노드가 다수의 작업 노드 각각이 동일 클러스터 내의 다른 작업 노드가 빅데이터를 처리한 결과를 분석하여 생성한 로컬 신뢰 의견을 수집하고, 수집된 로컬 신뢰 의견 사이의 유사도를 기반으로 주관 평판을 계산하여, 클러스터 내의 작업 노드 중 유사한 주관 평판을 갖는 과반 이상의 작업 노드를 그룹화하여 신뢰 정족수 집합을 획득하고, 획득된 상기 신뢰 정족수 집합 내의 작업 노드의 주관 평판을 기반으로 각 작업 노드가 다른 작업 노드의 빅데이터 처리한 행동을 평가한 행동 평판을 계산하며, 상기 로컬 신뢰 의견과 상기 행동 평판을 기반으로 각 작업 노드가 다른 작업 노드를 평가하는 행동에 대한 신뢰도 평판을 계산하여 획득하는 글로벌 평판인 빅데이터 처리 제어 방법.

청구항 17

제16항에 있어서, 상기 로컬 신뢰 의견(LTO)은 수학적식

$$LTO_{j,k} = \frac{p_{j,k}}{p_{j,k} + n_{j,k}}$$

(여기서 $LTO_{j,k}$ 는 클러스터 내의 제j 작업 노드(WN_j)가 제k 작업 노드(WN_k)에 대해 측정한 로컬 신뢰 의견, $p_{j,k}$ 는 기지정된 시간 동안 관측된 긍정적 행동 횟수로 정상적으로 빅데이터를 처리한 횟수이고, $n_{j,k}$ 는 부정적 행동 횟수로 빅데이터 처리에 요구되는 목표 시간(T_{object})을 초과하거나, 통신 오류가 발생하는 등의 비정상적인 행동이 발생한 횟수이다.)

상기 주관 평판(SR)은

$$SR_{w,u} = \sum_{j \in S_u} LTO_{j,u} \frac{HR_j \text{sim}(w,j)}{\sum_{j \in S_u} HR_j \text{sim}(w,j)}$$

(여기서 S_u 는 클러스터 내의 제u 작업 노드(WN_u)에 대한 로컬 신뢰 의견(LTO)을 전송한 작업 노드 집합을 나타내고, $\text{sim}(w,j)$ 는 제w 작업 노드(WN_w)와 제j 작업 노드(WN_j)가 보고한 로컬 신뢰 의견(LTO) 사이의 코사인 유사도를 나타내며, HR_j 는 로컬 신뢰 의견(LTO)을 생성한 각 작업 노드의 계층별 가중치로서 일반 작업 노드(WN)는 1이고, 클러스터 헤드 노드(CH)는 1보다 큰 기지정된 값을 갖는다.)

에 따라 계산되며,

상기 행동 평판(BR)은 수학적식

$$BR_{CH_i,u} = \frac{\sum_{w \in D} SR_{w,u}}{|D_{CH_i}|}$$

(여기서 $|D_{CH_i}|$ 는 제i 클러스터(CL_i)의 클러스터 헤드(CH_i)가 구성한 신뢰 정족수 집합(D)에 포함된 작업 노드의 개수를 나타낸다.)

에 따라 계산되고,

상기 신뢰도 평판(CR)은 수학적식

$$CR_{CH_i,u} = 1 - \sqrt{\frac{\sum_{j \in \{LTO_{u,j} \neq null\}} (LTO_{u,j} - BR_{CH_i,u})^2}{|\{j \in \{LTO_{u,j} \neq null\}\}|}}$$

에 따라 계산되는 빅데이터 처리 제어 방법.

청구항 18

제15항에 있어서, 상기 최적 작업 노드 개수를 계산하는 단계는

상기 마스터 처리 시간을 상기 빅데이터 처리 제어 장치와 상기 클러스터 헤드 노드 사이의 변수 공유 시간과

데이터 전송 시간 및 데이터 전송 오류를 복구하기 위한 오류 복구 시간의 합으로 계산하는 단계;

상기 클러스터 처리 시간을 상기 클러스터 헤드 노드와 동일 클러스터 내의 나머지 작업 노드 사이의 변수 공유 시간과 데이터 전송 시간, 데이터 전송 오류 및 작업 노드들의 빅데이터 처리 오류를 복구하기 위한 오류 복구 시간 및 데이터 처리 시간의 합으로 계산하는 단계; 및

상기 마스터 처리 시간과 상기 클러스터 처리 시간의 합으로 계산되는 상기 빅데이터 처리 시간이 상기 목표 시간 이하가 되도록 하는 최소값을 최적 작업 노드 개수로서 획득하는 단계를 포함하는 빅데이터 처리 제어 방법.

청구항 19

제18항에 있어서, 상기 처리 노드 수를 판별하는 단계는

상기 정적 작업 노드의 개수가 상기 최적 작업 노드 개수 이상이면, 상기 최적 작업 노드 개수를 상기 처리 노드 수로 결정하는 단계; 및

상기 정적 작업 노드의 개수가 상기 최적 작업 노드 개수보다 작으면, 상기 정적 작업 노드의 개수를 상기 처리 노드 수로 결정하는 단계를 포함하는 빅데이터 처리 제어 방법.

청구항 20

제11항에 있어서, 상기 분배하는 단계는

데이터를 인가받아 미리 수집된 빅데이터를 상기 처리 노드 수에 따라 분할하고, 분할된 빅데이터를 각 클러스터에 포함된 정적 작업 노드의 개수를 고려하여 클러스터 단위로 분배하는 단계; 및

각 클러스터에서 처리된 빅데이터를 인가받아 수집하여 작업 명령을 인가한 사용자 단말로 전송하는 단계를 포함하는 빅데이터 처리 제어 방법.

발명의 설명

기술 분야

[0001] 본 발명은 빅데이터 처리 제어 장치 및 방법에 관한 것으로, 적응적 클러스터링을 이용하여 보안성을 향상시키면서 실시간으로 빅데이터를 처리할 수 있는 빅데이터 처리 제어 장치 및 방법에 관한 것이다.

배경 기술

[0002] 사물인터넷(IoT), 드론, 자율 주행 차량, 확장현실(eXtended Reality, XR) 등 다양한 대화형 영상 서비스 및 시스템이 크게 증가함에 따라, 실시간으로 처리해야하는 데이터의 양은 날이 갈수록 증가하고 있다. 그리고 기존의 오프라인 서버를 활용한 데이터 센터만으로는 이와 같이 수집되는 방대한 양의 빅데이터들이 요구하는 데이터 처리 성능을 보장하지 못한다. 따라서 증가된 데이터 처리 수요를 충족시키기 위해 인터넷 및 클라우드 등을 통해 물리적으로 구분되어 있는 서버들을 연결하여 빅데이터를 처리하려는 시도가 생겨나고 있다.

[0003] 빅데이터 시스템은 대량의 정형 및 비정형 데이터를 수집하기 때문에, 다양한 네트워크 서버 및 클라우드와 상호 연결되며, 따라서 빅데이터 시스템은 다양한 형태의 인터페이스를 갖추어야 한다. 빅데이터 시스템이 다양한 인터페이스를 갖추면 접근성이 높아지는 반면, 해당 인터페이스를 통한 보안 공격의 가능성도 동시에 높아지게 되는 문제를 야기한다.

[0004] 이러한 문제를 해결하기 위해, 빅데이터 시스템의 보안성을 향상시키기 위한 여러 기법들이 제시되어 왔고, 이 중에는 신뢰 기반 보안 기법을 활용한 빅데이터 시스템의 보안 기법들도 제안되었다. 다만 기존의 신뢰 기반 보안 기법은 대부분 데이터를 수집하는 센서나 사물인터넷 기기들 또는 수집된 데이터를 운반하는 드론 및 자율 주행차량 등에 대한 신뢰도를 측정함으로써, 빅데이터 시스템의 보안을 유지하고자 하였다. 그러나 해당 기법들은 빅데이터를 처리하는 작업 노드들의 신뢰도는 고려하지 않았다. 이로 인해 작업 노드를 대상으로 하는 보안 공격에 취약하다는 한계가 있었다. 또한 보안성 향상을 위한 절차를 수행하는 경우, 보안 절차로 인해 인가된 데이터에 대한 응답 속도가 저하되어 실시간성을 요구하는 작업에 적용될 수 없다. 즉 작업이 요구하는 목표 시간을 만족시킬 수 없다는 문제가 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국 등록 특허 제10-2260094호 (2021.05.28 등록)

발명의 내용**해결하려는 과제**

[0006] 본 발명의 목적은 다수의 작업 노드들을 클러스터링하고 신뢰도에 따라 빅데이터를 분배하여 병렬 처리함으로써, 보안성을 향상시키면서도 작업 명령이 요구하는 목표 시간을 만족시킬 수 있는 빅데이터 처리 제어 장치 및 방법을 제공하는데 있다.

[0007] 본 발명의 다른 목적은 다수의 작업 노드들에 대한 신뢰도를 계산하는 계산 복잡도가 최소화되도록 클러스터링하여 신뢰도 계산에 따른 응답 지연을 최소로 억제할 수 있는 빅데이터 처리 제어 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0008] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 빅데이터 처리 제어 장치는 상기 다수의 작업 노드에 대한 신뢰도를 계산하는 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하여 상기 다수의 작업 노드를 계산된 최적 클러스터 개수의 클러스터로 구분하고, 상기 다수의 작업 노드 각각에 대해 획득된 신뢰도 정보를 기반으로 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 판별하며, 상기 작업 명령에서 요구하는 시간인 목표 시간 이내에 빅데이터를 구분된 클러스터로 분산 처리할 수 있는 작업 노드의 개수인 최적 작업 노드 개수를 계산하고, 상기 정직 작업 노드의 개수와 상기 최적 작업 노드 개수에 따라 실제 빅데이터를 분산처리하는 작업 노드의 개수인 처리 노드 수를 결정하는 시스템 관리부; 및 데이터를 인가받아 수집하여 빅데이터를 생성하고, 생성된 빅데이터를 상기 처리 노드 수의 작업 노드가 처리하도록 분배하는 빅데이터 관리부를 포함한다.

[0009] 상기 시스템 관리부는 구분되는 클러스터의 개수에 따른 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하고, 계산된 최적 클러스터 개수에 따라 상기 다수의 작업 노드를 클러스터링하여 구분하며, 구분된 각 클러스터에 포함된 작업 노드 각각의 신뢰도에 따라 클러스터를 관리하는 클러스터 헤드 노드를 지정하는 클러스터 설정부; 구분된 각 클러스터의 클러스터 헤드 노드가 대응하는 클러스터 내의 작업 노드 각각에 대해 획득한 신뢰도 정보를 인가받아, 상기 다수의 작업 노드에서 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 확인하는 신뢰도 기반 판별부; 상기 빅데이터 처리 제어 장치가 상기 빅데이터를 각 클러스터로 분배하기 위해 소요되는 마스터 처리 시간과 각 클러스터가 분배된 빅데이터를 처리하기 위해 소요되는 클러스터 처리 시간을 계산하여 상기 작업 명령에 대응하여 빅데이터를 처리하는 빅데이터 처리 시간을 획득하고, 상기 빅데이터 처리 시간이 상기 목표 시간을 만족시킬 수 있도록 하는 최소의 작업 노드 개수인 최적 작업 노드 개수를 계산하는 최적 개수 판별부; 및 상기 정직 작업 노드의 개수와 상기 최적 작업 노드 개수를 비교하여 상기 처리 노드 수를 결정하는 작업 노드수 결정부를 포함할 수 있다.

[0010] 상기 클러스터 설정부는 상기 빅데이터 시스템의 환경 변화를 분석하여 클러스터 재구성 여부를 판별하는 클러스터 재구성 판별부; 클러스터를 재구성해야 하는 것으로 판별되면, 전체 작업 노드의 개수에 따라 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하는 최적 클러스터 결정부; 계산된 최적 클러스터 개수에 따라 작업 노드를 다수의 작업 노드를 클러스터링하여 구분하는 작업 노드 클러스터링부; 및 각 클러스터에 포함된 작업 노드 각각의 이전 획득되어 저장된 신뢰도 정보를 확인하여, 가장 높은 신뢰도를 갖는 작업 노드를 해당 클러스터의 클러스터 헤드 노드로 선택하는 클러스터 헤드 선택부를 포함할 수 있다.

[0011] 상기 신뢰도 기반 판별부는 구분된 각 클러스터의 클러스터 헤드 노드가 대응하는 클러스터 내의 작업 노드 각각에 대해 획득한 신뢰도 정보를 인가받아 기지정된 방식으로 검증하는 신뢰도 정보 검증부; 검증된 신뢰도 정보에 따라 상기 다수의 작업 노드 각각을 상기 정직 작업 노드 또는 상기 악의적 작업 노드로 판별하는 악의적 노드 선별부; 및 판별된 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 확인하는 신뢰 노드수 확인부를 포함할 수 있다.

[0012] 상기 신뢰도 정보는 각 클러스터에서 상기 클러스터 헤드 노드가 다수의 작업 노드 각각이 동일 클러스터 내의

다른 작업 노드가 빅데이터를 처리한 결과를 분석하여 생성한 로컬 신뢰 의견을 수집하고, 수집된 로컬 신뢰 의견 사이의 유사도를 기반으로 주관 평판을 계산하여, 클러스터 내의 작업 노드 중 유사한 주관 평판을 갖는 과반 이상의 작업 노드를 그룹화하여 신뢰 정족수 집합을 획득하고, 획득된 상기 신뢰 정족수 집합 내의 작업 노드의 주관 평판을 기반으로 각 작업 노드가 다른 작업 노드의 빅데이터 처리한 행동을 평가한 행동 평판을 계산하며, 상기 로컬 신뢰 의견과 상기 행동 평판을 기반으로 각 작업 노드가 다른 작업 노드를 평가하는 행동에 대한 신뢰도 평판을 계산하여 획득하는 글로벌 평판일 수 있다.

[0013] 상기 최적 개수 판별부는 상기 마스터 처리 시간을 상기 빅데이터 처리 제어 장치와 상기 클러스터 헤드 노드 사이의 변수 공유 시간과 데이터 전송 시간 및 데이터 전송 오류를 복구하기 위한 오류 복구 시간의 합으로 계산하는 전송 시간 계산부; 상기 클러스터 처리 시간을 상기 클러스터 헤드 노드와 동일 클러스터 내의 나머지 작업 노드 사이의 변수 공유 시간과 데이터 전송 시간, 데이터 전송 오류 및 작업 노드들의 빅데이터 처리 오류를 복구하기 위한 오류 복구 시간 및 데이터 처리 시간의 합으로 계산하는 작업 시간 계산부; 및 상기 마스터 처리 시간과 상기 클러스터 처리 시간의 합으로 계산되는 상기 빅데이터 처리 시간이 상기 목표 시간 이하가 되도록 하는 최소값을 최적 작업 노드 개수로서 획득하는 최적 개수 계산부를 포함할 수 있다.

[0014] 상기 작업 노드수 결정부는 상기 정직 작업 노드의 개수가 상기 최적 작업 노드 개수 이상이면, 상기 최적 작업 노드 개수를 상기 처리 노드 수로 결정하고, 상기 정직 작업 노드의 개수가 상기 최적 작업 노드 개수보다 작으면, 상기 정직 작업 노드의 개수를 상기 처리 노드 수로 결정할 수 있다.

[0015] 상기 빅데이터 관리부는 데이터를 인가받아 수집하여 빅데이터를 생성하고, 각 클러스터에서 처리된 빅데이터를 인가받아 수집하여 작업 명령을 인가한 사용자 단말로 전송하는 빅데이터 수집부; 및 빅데이터를 상기 처리 노드 수에 따라 분할하고, 분할된 빅데이터를 각 클러스터에 포함된 정직 작업 노드의 개수를 고려하여 클러스터 단위로 분배하는 빅데이터 분배부를 포함할 수 있다.

[0016] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 빅데이터 처리 제어 방법은 작업 명령에 따라 처리되어야 하는 빅데이터를 분산 처리하는 다수의 작업 노드에 대한 신뢰도를 계산하는 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수를 계산하여 상기 다수의 작업 노드를 계산된 최적 클러스터 개수의 클러스터로 구분하고, 상기 다수의 작업 노드 각각에 대해 획득된 신뢰도 정보를 기반으로 정직 작업 노드의 개수와 악의적 작업 노드의 개수를 판별하며, 상기 작업 명령에서 요구하는 시간인 목표 시간 이내에 빅데이터를 구분된 클러스터로 분산 처리할 수 있는 작업 노드의 개수인 최적 작업 노드 개수를 계산하고, 상기 정직 작업 노드의 개수와 상기 최적 작업 노드 개수에 따라 실제 빅데이터를 분산처리하는 작업 노드의 개수인 처리 노드 수를 결정하는 단계; 및 데이터를 인가받아 수집하여 빅데이터를 생성하고, 생성된 빅데이터를 상기 처리 노드 수의 작업 노드가 처리하도록 분배하는 단계를 포함한다.

발명의 효과

[0017] 따라서, 본 발명의 실시예에 따른 빅데이터 처리 제어 장치 및 방법은 작업 노드들에 대한 신뢰도를 판정하는 계산 복잡도가 최소화되도록 신뢰도에 기반하여 다수의 작업 노드들을 클러스터링하여 빅데이터를 분산 병렬 처리함으로써, 보안성을 향상시키면서 작업 명령이 요구하는 목표 시간을 만족시킬 수 있도록 하며, 악의적 작업 노드를 빅데이터 처리에서 배제시킴으로써 작업 노드에 대한 보안 공격 상황에서도 빅데이터 처리 성능이 저하되는 것을 가능한 억제할 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 일 실시예에 따른 빅데이터 시스템의 개략적 구조를 나타낸다.
 도 2는 도 1의 마스터 노드의 개략적 구성의 일 예를 나타낸다.
 도 3은 도 1의 작업 노드의 개략적 구성의 일 예를 나타낸다.
 도 4는 도 2의 클러스터 설정부의 상세 구성의 일 예를 나타낸다.
 도 5는 작업 노드 개수와 클러스터 개수에 따른 신뢰도 계산 복잡도 관계를 나타내는 그래프이다.
 도 6은 마스터 노드가 다수의 작업 노드 중 클러스터 헤드 노드를 지정하기 위해 전송하는 메시지의 일 예를 나타낸다.
 도 7은 도 2의 신뢰도 기반 판별부의 상세 구성의 일 예를 나타낸다.

도 8은 도 2의 최적 개수 판별부의 상세 구성의 일 예를 나타낸다.

도 9는 본 발명의 일 실시예에 따른 빅데이터 처리 제어 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0019] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.
- [0020] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.
- [0021] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0022] 도 1은 본 발명의 일 실시예에 따른 빅데이터 시스템의 개략적 구조를 나타낸다.
- [0023] 도 1을 참조하면, 본 실시예에 따른 빅데이터 시스템은 마스터 노드(MN)와 신뢰도 관리 데이터베이스(DB) 및 다수의 작업 노드(WN)를 포함한다. 여기서 마스터 노드(MN)는 빅데이터 시스템에서 빅데이터 처리 제어 장치로서 동작한다.
- [0024] 마스터 노드(MN)는 IoT 센서나 드론, 자율 주행 차량 등과 같은 각종 센서나 단말로부터 전송되는 빅데이터를 인가받아 관리하고, 사용자 단말(User)로부터 빅데이터를 처리하기 위한 작업 명령이 인가되면, 인가된 작업 명령에 따라 처리해야할 빅데이터를 다수의 작업 노드(WN)로 분배한다. 그리고 다수의 작업 노드(WN)가 분배된 빅데이터를 처리하여 전송하면, 처리된 빅데이터를 작업 결과로서 사용자 단말(User)로 전달한다. 다만 마스터 노드(MN)는 직접 다수의 작업 노드(WN)로 빅데이터를 분배하는 않고, 다수의 작업 노드(WN)를 적어도 하나의 클러스터($CL_1 \sim CL_k$)로 구분하여 구분된 각 클러스터($CL_1 \sim CL_k$)의 클러스터 헤드 노드(CH)를 통해 빅데이터를 분배하며, 클러스터 헤드 노드(CH)를 통해 각 클러스터($CL_1 \sim CL_k$)에서 처리된 빅데이터를 전송받는다.
- [0025] 마스터 노드(MN)는 다수의 작업 노드(WN)를 적어도 하나의 클러스터($CL_1 \sim CL_k$)로 구분하기 위한 클러스터 개수를 우선 결정하고, 결정된 클러스터 개수에 따라 다수의 작업 노드(WN)를 클러스터링하여 적어도 하나의 클러스터($CL_1 \sim CL_k$)로 구분한다. 이때 본 실시예에 따른 마스터 노드(MN)는 다수의 작업 노드(WN)에 대한 신뢰도를 계산하는 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수(k_{opt})를 계산하고, 계산된 최적 클러스터 개수(k_{opt})에 따라 다수의 작업 노드(WN)를 구분한다.
- [0026] 여기서 마스터 노드(MN)가 다수의 작업 노드(WN)에 대한 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수(k_{opt})를 계산하는 것은, 작업 노드(WN)의 개수가 증가될수록 빅데이터는 분산 처리되어 처리 시간이 짧아지는 반면, 다수의 작업 노드(WN)에 대한 신뢰도 계산 시간은 오히려 증가되어 실시간 처리가 요구되는 작업의 목표 시간(T_{object})을 도달할 수 없도록 하기 때문이다.
- [0027] 신뢰도 계산은 다수의 작업 노드(WN)는 상호 신뢰도를 우선 계산하고, 이후 마스터 노드(MN)가 각 작업 노드가 계산한 신뢰도를 검증하는 방식으로 수행된다. 이때, 다수의 작업 노드(WN)를 클러스터($CL_1 \sim CL_k$)로 구분하고, 구분된 클러스터 내의 작업 노드(WN)가 상호 신뢰도를 계산하면 신뢰도 계산량을 줄일 수 있으므로 계산 복잡도가 낮아진다. 다만 클러스터($CL_1 \sim CL_k$)의 개수가 증가할수록, 마스터 노드(MN)에서 수행되는 신뢰도 검증에 대한 계산량이 증가하게 된다. 따라서 신뢰도 계산에 의한 응답 지연을 억제하기 위해 클러스터($CL_1 \sim CL_k$)의 개수는 각 클러스터에서 수행되는 신뢰도 계산과 마스터 노드(MN)에서 수행되는 검증 계산을 함께 고려하여 계산 복잡도가 최소화되는 개수(k_{opt})로 결정되어야 한다.
- [0028] 즉 본 실시예에서 마스터 노드(MN)는 작업 노드(WN)의 개수의 증가에 따라 신뢰도 계산을 위한 지연 시간으로 인해 실시간 작업을 처리할 수 없게 되는 문제가 발생되지 않도록 빅데이터 시스템에 포함된 작업 노드(WN)의

전체 개수(n)에 따른 최적 클러스터 개수(k_{opt})를 계산하고, 계산된 최적 클러스터 개수(k_{opt})로 다수의 작업 노드(WN)를 구분한다.

[0029] 마스터 노드(MN)는 초기 구동 시 또는 기지정된 주기로 최적 클러스터 개수(k_{opt})를 재계산하여 다수의 작업 노드(WN)를 적어도 하나의 클러스터($CL_1 \sim CL_k$)로 구분할 수 있다. 또한 마스터 노드(MN)는 작업 명령 변경이나 작업 노드(WN)의 상태 변경 등과 같은 빅데이터 시스템의 환경 변화를 추적하고 분석하여, 환경 변화가 발생하는 경우에도 최적 클러스터 개수(k_{opt})를 재계산할 수 있다.

[0030] 그리고 마스터 노드(MN)는 최적 클러스터 개수(k_{opt})로 구분된 클러스터($CL_1 \sim CL_k$) 각각에 포함된 작업 노드(WN)들을 관리하기 위한 클러스터 헤드 노드(CH)를 선택한다. 이때 마스터 노드(MN)는 신뢰도 관리 데이터베이스(DB)에 이전 저장된 각 작업 노드(WN)의 신뢰도 정보를 기반으로 각 클러스터($CL_1 \sim CL_k$)의 클러스터 헤드 노드(CH)를 선택할 수 있다.

[0031] 한편 마스터 노드(MN)는 빅데이터 시스템의 보안성을 향상시키기 위해, 각 클러스터($CL_1 \sim CL_k$)의 클러스터 헤드 노드(CH)가 대응하는 클러스터의 작업 노드(WN)로부터 수집한 다수의 작업 노드(WN) 각각에 대한 신뢰도 정보를 인가받아 검증하고, 검증된 다수의 작업 노드(WN)에 대한 신뢰도 정보를 신뢰도 관리 데이터베이스(DB)로 전송한다. 그리고 검증된 신뢰도 정보를 기반으로 다수의 작업 노드(WN) 중 신뢰할 수 있는 정직 작업 노드(WN^H)와 악의적 작업 노드(WN^M)를 판별하여, 정직 작업 노드(WN^H)의 개수(N^H)와 악의적 작업 노드(WN^M)의 개수(N^M)를 확인한다. 이와 함께 마스터 노드(MN)는 사용자 단말(User)에서 빅데이터를 이용하여 처리하도록 요청된 작업 명령이 요청하는 목표 시간(T_{object})을 만족시킬 수 있는 최소 작업 노드 개수를 나타내는 최적 작업 노드 개수(N_{opt})를 판별한다.

[0032] 정직 작업 노드(WN^H)의 개수(N^H)와 최적 작업 노드 개수(N_{opt})가 판별되면, 마스터 노드(MN)는 판별된 정직 작업 노드(WN^H)의 개수(N^H)와 최적 작업 노드 개수(N_{opt})에 따라 빅데이터를 처리할 작업 노드(WN)의 개수인 처리 노드 수(N_{pro})를 결정한다. 그리고 마스터 노드(MN)는 결정된 처리 노드 수(N_{pro})에 따라 작업 명령에 따라 처리해야 할 빅데이터를 분할하여 각 클러스터($CL_1 \sim CL_k$)로 전달한다. 마스터 노드(MN)는 분할된 빅데이터를 구분된 클러스터 각각에서 지정된 클러스터 헤드 노드(CH)로 전달하여, 각 클러스터 헤드 노드(CH)가 대응하는 클러스터 내에 포함된 작업 노드들로 분배되어 전달된 빅데이터를 다시 분배할 수 있도록 한다. 즉 마스터 노드(MN)는 인가된 빅데이터가 다수의 작업 노드(WN)에서 병렬로 분산처리되도록 함으로써 빅데이터 처리 속도를 향상시킨다. 그리고 마스터 노드(MN)는 각 클러스터($CL_1 \sim CL_k$)의 다수의 작업 노드(WN)에서 분산 처리된 빅데이터를 클러스터 헤드 노드(CH)를 통해 인가받아 사용자 단말(User)로 전달한다.

[0033] 마스터 노드(MN)는 다수의 작업 노드(WN) 중 악의적 작업 노드(WN^M)를 배제하고 나머지 정직 작업 노드(WN^H) 중 결정된 처리 노드 수(N_{pro})의 작업 노드(WN)가 빅데이터를 처리하도록 하여 작업 노드(WN)에 대한 보안 공격에도 빅데이터 처리 시스템의 보안성이 유지되도록 할 수 있다. 이를 위해 마스터 노드(MN)는 각 클러스터($CL_1 \sim CL_k$)에서 악의적 노드(WN^M)를 제외한 정직 작업 노드(WN^H)의 수를 확인하고, 확인된 각 클러스터별 정직 작업 노드(WN^H)의 수에 기반하여 빅데이터를 분배할 수 있다.

[0034] 신뢰도 관리 데이터베이스(DB)는 마스터 노드(MN)에서 검증된 다수의 작업 노드(WN) 각각에 대한 신뢰도 정보를 인가받아 저장하고, 마스터 노드(MN)의 요청에 따라 저장된 신뢰도 정보를 전송한다. 여기서는 신뢰도 관리 데이터베이스(DB)를 마스터 노드(MN)와 구분하여 도시하였으나, 신뢰도 관리 데이터베이스(DB)는 마스터 노드(MN)에 포함되어 구성될 수도 있다.

[0035] 다수의 작업 노드(WN) 각각은 마스터 노드(MN)에 의해 다수의 클러스터($CL_1 \sim CL_k$) 중 하나의 클러스터에 포함되도록 구분되고, 마스터 노드(MN)에서 분배되어 전송되는 빅데이터를 인가받아 처리한다. 이때 구분된 클러스터($CL_1 \sim CL_k$) 각각에서 다수의 작업 노드(WN) 중 하나는 마스터 노드(MN)에 의해 클러스터 헤드 노드(CH)로 선택되며, 선택된 클러스터 헤드 노드(CH)는 마스터 노드(MN)가 클러스터($CL_1 \sim CL_k$)에 따라 분배한 빅데이터를 인가

받고, 인가된 빅데이터를 다시 대응하는 클러스터 내부에 포함된 작업 노드(WN)로 분배한다. 클러스터 헤드 노드(CH)는 악의적 작업 노드(WN^M)를 제외한 정직 작업 노드(WN^H)로 빅데이터를 분배할 수 있다.

[0036] 결과적으로 본 실시예에서 빅데이터 처리 제어 장치로 동작하는 마스터 노드는 빅데이터 시스템에서 작업 노드(WN)의 전체 개수(n)에 따라 신뢰도 계산 복잡도를 최소화할 수 있는 최적 클러스터 개수(k_{opt})를 계산하여 다수의 작업 노드를 클러스터링함으로써 신뢰도 계산에 따른 응답 지연을 최대한 억제하면서 빅데이터가 분산 병렬 처리되도록 한다. 또한 다수의 작업 노드(WN) 각각에 대해 계산된 신뢰도에 따라 판별된 악의적 작업 노드(WN^M)를 배제하면서 빅데이터를 처리하기 위한 최적의 작업 노드(WN) 개수인 처리 노드 수(N_{pro})를 도출하여, 보안을 향상시키면서도 작업 명령에 요구되는 목표 시간(T_{object})을 만족시킬 수 있도록 한다.

[0037] 도 2는 도 1의 마스터 노드의 개략적 구성의 일 예를 나타낸다.

[0038] 도 2를 참조하면, 마스터 노드(MN)는 통신부(110), 시스템 관리부 및 빅데이터 관리부를 포함할 수 있다.

[0039] 통신부(110)는 다양한 센서 또는 단말 등의 각종 장치들로부터 데이터를 인가받는다. 그리고 사용자 단말(User)로부터 빅데이터를 처리하기 위한 작업 명령을 인가받거나, 처리된 빅데이터를 사용자 단말(User)로 전송한다. 또한 통신부(110)는 미리 결정된 클러스터(CL₁ ~ CL_k) 각각의 클러스터 헤드 노드(CH)로 분배된 빅데이터로 전송하거나, 각 클러스터(CL₁ ~ CL_k)에서 처리된 빅데이터를 인가받으며, 각 클러스터(CL₁ ~ CL_k)에 포함된 다수의 작업 노드(WN) 각각에 대한 신뢰도 정보를 인가받는다. 그리고 신뢰도 관리 데이터베이스(DB)로 다수의 작업 노드(WN)에 대한 신뢰도 정보를 전달하거나, 신뢰도 관리 데이터베이스(DB)로부터 신뢰도 정보를 인가받는다.

[0040] 시스템 관리부는 다수의 작업 노드(WN)를 다수의 클러스터(CL₁ ~ CL_k)로 구분하며, 다수의 작업 노드(WN) 각각을 신뢰도에 따라 정직 작업 노드(WN^H)와 악의적 작업 노드(WN^M)를 구분하고, 작업 명령이 요구하는 목표 시간(T_{object})을 만족할 수 있는 최적 작업 노드 수(N_{opt})를 계산하여, 빅데이터를 처리할 작업 노드(WN)의 개수인 처리 노드 수(N_{pro})를 결정한다.

[0041] 시스템 관리부는 클러스터 설정부(120), 신뢰도 기반 판별부(130), 최적 개수 판별부(140) 및 작업 노드수 결정부(150)를 포함할 수 있다.

[0042] 클러스터 설정부(120)는 빅데이터 시스템의 환경을 분석하여 전체 작업 노드(WN)의 개수(n)에 대응하는 최적 클러스터 개수(k_{opt})를 계산하고, 계산된 최적 클러스터 개수(k_{opt})에 따라 다수의 작업 노드(WN)를 클러스터링하여 최적 클러스터 개수(k_{opt})의 클러스터(CL₁ ~ CL_k)로 구분한다.

[0043] 클러스터 설정부(120)는 마스터 노드(MN)의 초기 구동 시와 기지정된 주기로 최적 클러스터 개수(k_{opt})를 계산하여 클러스터링을 수행할 수 있으며, 빅데이터 시스템의 환경 요인에 변화가 발생된 경우에도 최적 클러스터 개수(k_{opt})를 재계산하여 클러스터링을 수행할 수 있다. 빅데이터 시스템에서는 다수의 작업 노드(WN)의 위치, 성능, 네트워크 인터페이스 등이 서로 상이하므로, 새로운 작업 노드(WN)가 시스템에 추가되거나 기존의 작업 노드(WN)가 시스템에서 제거되는 경우 및 이전 정직 작업 노드(WN^H)로 판단된 작업 노드가 보안 공격 등으로 인해 악의적 작업 노드(WN^M)으로 변경된 경우에도 최적 클러스터 개수(k_{opt})를 재계산하여 클러스터링을 수행할 수 있다. 그리고 빅데이터를 처리하기 위한 요청된 작업 명령이 수행되어야 하는 목표 시간(T_{object})이 서로 상이할 수 있으므로, 마스터 노드(MN)는 작업 명령이 변경되는 경우에 최적 클러스터 개수(k_{opt})를 재계산할 수도 있다.

[0044] 추가적으로 클러스터 설정부(120)는 구분된 적어도 하나의 클러스터(CL₁ ~ CL_k) 각각의 다수의 작업 노드(WN) 중 해당 클러스터를 관리할 클러스터 헤드 노드(CH)를 선택한다.

[0045] 신뢰도 기반 판별부(130)는 적어도 하나의 클러스터(CL₁ ~ CL_k) 각각의 클러스터 헤드 노드(CH)로부터 각 클러스터(CL₁ ~ CL_k)에 포함된 작업 노드(WN)에 대한 신뢰도 정보를 인가받아 검증하여, 검증된 신뢰도 정보를 기반으로 정직 작업 노드(WN^H)와 악의적 작업 노드(WN^M)를 판별하고, 판별된 정직 작업 노드(WN^H)의 개수(N^H)와 악의적

작업 노드(WN^M)의 개수(N^M)를 확인한다.

- [0046] 최적 개수 판별부(140)는 각 클러스터($CL_1 \sim CL_k$)에서 분배된 빅데이터를 처리하기 위해 소요되는 클러스터 처리 시간(T_c^k)과 마스터 노드(MN)에서 빅데이터를 처리하기 위해 소요되는 마스터 처리 시간(T_m)을 기반으로 작업 명령에서 요청된 목표 시간(T_{object})을 만족시키기 위한 최적 작업 노드 개수(N_{opt})를 판별한다.
- [0047] 작업 노드수 결정부(150)는 신뢰도 기반 판별부(130)에서 판별된 정직 작업 노드(WN^H)의 개수(N^H)와 악의적 작업 노드(WN^M)의 개수(N^M)와 최적 개수 판별부(140)에서 판별된 최적 작업 노드 개수(N_{opt})를 기반으로 작업 명령에 따라 빅데이터를 처리할 작업 노드(WN)의 개수로 처리 노드 수(N_{pro})를 결정한다. 이때 작업 노드수 결정부(150)는 각 클러스터($CL_1 \sim CL_k$)에서 악의적 작업 노드(WN^M)를 배제한 정직 작업 노드(WN^H)에 기반하여 빅데이터를 처리할 작업 노드(W)의 개수를 함께 결정할 수 있다.
- [0048] 한편 빅데이터 관리부는 통신부(110)를 통해 전송된 데이터로 빅데이터를 구성하여 저장하고, 저장된 빅데이터 중 작업 명령에 따라 처리되어야 하는 빅데이터를 시스템 관리부에서 결정된 처리 노드 수(N_{pro})로 분할하여 클러스터($CL_1 \sim CL_k$) 각각의 클러스터 헤드 노드(CH)로 분배한다.
- [0049] 빅데이터 관리부는 빅데이터 수집부(160) 및 빅데이터 분배부(170)를 포함할 수 있다.
- [0050] 빅데이터 수집부(160)는 통신부(110)를 통해 다수의 센서 또는 단말에서 전송된 데이터를 인가받아 빅데이터를 구성하여 저장한다. 그리고 빅데이터 수집부(160)는 사용자 단말(User)에서 전송된 작업 명령에 따라 처리되어야 하는 빅데이터를 빅데이터 분배부(170)로 전달한다. 빅데이터 수집부(160)는 빅데이터를 관리하는 데이터 관리 데이터베이스로서, 경우에 따라서는 별도의 장치로 구분될 수 있다.
- [0051] 빅데이터 분배부(170)는 빅데이터 수집부(160)로부터 작업 명령에 따라 처리되어야 하는 빅데이터를 인가받고, 시스템 관리부에서 결정된 처리 노드 수(N_{pro})에 따라 인가된 빅데이터를 분할한다. 즉 처리되어야 하는 빅데이터를 N_{pro} 개로 분할한다. 그리고 각 클러스터($CL_1 \sim CL_k$)에서 빅데이터를 처리하도록 지정된 작업 노드(WN)의 개수에 따라 분할된 빅데이터를 각 클러스터($CL_1 \sim CL_k$)의 클러스터 헤드 노드(CH)로 분배한다.
- [0052] 그리고 빅데이터 분배부(170)는 각 클러스터($CL_1 \sim CL_k$)에 분배되어 처리된 빅데이터를 클러스터 헤드 노드(CH)를 통해 인가받아 빅데이터 수집부(160)로 전달하고, 빅데이터 수집부(160)는 처리된 빅데이터를 취합하여 사용자 단말(User)로 전달한다.
- [0053] 도 2에서 클러스터 설정부(120)와 신뢰도 기반 판별부(130) 및 최적 개수 판별부(140)의 상세 구성과 동작은 후술하도록 한다.
- [0054] 도 3은 도 1의 작업 노드의 개략적 구성의 일 예를 나타낸다.
- [0055] 도 3을 참조하면, 다수의 작업 노드(WN) 각각은 작업 통신부(210), 데이터 처리부(220), 로컬 신뢰 의견 생성부(230), 데이터 분배부(240) 및 클러스터 신뢰도 관리부(250)를 포함할 수 있다. 여기서 작업 통신부(210)와 데이터 처리부(220) 및 로컬 신뢰 의견 생성부(230)는 빅데이터 시스템의 모든 작업 노드(WN)에서 수행되어야 하는 동작을 위해 구비되며, 데이터 분배부(240) 및 클러스터 신뢰도 관리부(250)는 작업 노드(WN)가 해당 클러스터의 클러스터 헤드 노드(CH)로 선택된 경우, 클러스터 헤드 노드(CH)로서 동작하기 위해 구비된다. 다만 다수의 작업 노드(WN) 중 어느 작업 노드라도 마스터 노드(MN)에 의해 클러스터 헤드 노드(CH)로 선택될 수 있으므로, 다수의 작업 노드(WN)는 모두 작업 통신부(210)와 데이터 처리부(220) 및 로컬 신뢰 의견 생성부(230)뿐만 아니라 데이터 분배부(240)와 클러스터 신뢰도 관리부(250)를 포함하여 구성된다.
- [0056] 작업 통신부(210)는 해당 클러스터의 다른 작업 노드(WN)들과 통신을 수행한다. 그리고 작업 통신부(210)는 작업 노드(WN)가 클러스터 헤드 노드(CH)로 선택된 경우, 클러스터 내의 다른 작업 노드(WN) 뿐만 아니라 마스터 노드(MN)와도 통신을 수행한다.
- [0057] 데이터 처리부(220)는 클러스터 헤드 노드(CH)에서 분배되어 전달된 빅데이터를 작업 명령에 따라 처리한다.
- [0058] 로컬 신뢰 의견 생성부(230)는 기지정된 시간 동안 동일 클러스터 내의 다른 작업 노드(WN)가 빅데이터를 처리

하는 행동을 기반으로 로컬 신뢰 의견(local trust opinion: LTO)을 생성한다. 즉 로컬 신뢰 의견(LTO)은 각 클러스터($CL_1 \sim CL_k$) 내의 다수의 작업 노드(WN)가 상호 관측을 통해 획득한 신뢰도로서, 신뢰도 계산을 위한 기초 정보로 이용된다. 클러스터 내의 다수의 작업 노드 중 제j 작업 노드(WN_j)가 제k 작업 노드(WN_k)에 대해 측정한 로컬 신뢰 의견($LTO_{j,k}$)은 수학적 식 1에 따라 계산될 수 있다.

수학적 식 1

$$LTO_{j,k} = \frac{p_{j,k}}{p_{j,k} + n_{j,k}}$$

여기서 $p_{j,k}$ 는 기지정된 시간 동안 관측된 긍정적 행동 횟수로 정상적으로 빅데이터를 처리한 횟수이고, $n_{j,k}$ 는 부정적 행동 횟수로 빅데이터 처리에 요구되는 목표 시간(T_{object})을 초과하거나, 통신 오류가 발생하는 등의 비정상적인 행동이 발생한 횟수이다.

각 작업 노드(WN)는 클러스터 내의 다른 모든 작업 노드(WN)에 대해 로컬 신뢰 의견(LTO)을 계산하고, 계산된 로컬 신뢰 의견(LTO)을 클러스터 헤드 노드(CH)로 전달한다.

데이터 분배부(240)는 작업 노드(WN)가 클러스터 헤드 노드(CH)로 선택되면 활성화되어 마스터 노드(MN)에서 클러스터에 따라 분배되어 전달된 빅데이터를 다시 클러스터 내의 작업 노드들로 분배한다. 데이터 분배부(240)는 마스터 노드(MN)에 의해 해당 클러스터에서 빅데이터를 처리하도록 결정된 개수의 작업 노드(WN)를 선택하고, 선택된 작업 노드(WN)로 빅데이터를 분배하여 전달한다. 이때 데이터 분배부(240)는 미리 관별된 악의적 작업 노드(WN^M)를 배제하고, 정직 작업 노드(WN^H)로만 빅데이터를 분배하여 전달한다. 여기서 악의적 작업 노드(WN^M)와 정직 작업 노드(WN^H)는 마스터 노드(MN)에서 관별되어 빅데이터와 함께 전달될 수 있으나, 경우에 따라서는 직접 악의적 작업 노드(WN^M)와 정직 작업 노드(WN^H)를 관별하도록 구성될 수도 있다.

그리고 데이터 분배부(240)는 클러스터 내의 다수의 작업 노드(WN)에서 처리된 빅데이터를 인가받아 마스터 노드(MN)으로 전달할 수 있다. 경우에 따라서 데이터 분배부(240)는 해당 클러스터 내에서 각 작업 노드(WN)가 빅데이터 처리 과정에서 발생한 오류를 확인하여 처리할 수도 있다.

클러스터 신뢰도 관리부(250)는 클러스터 내의 다수의 작업 노드(WN) 각각이 다른 작업 노드(WN)의 행동을 분석하여 수집한 로컬 신뢰 의견(LTO)을 인가받아, 클러스터 내의 다수의 작업 노드(WN) 각각에 대한 신뢰도 정보를 획득한다. 본 실시예에서 클러스터 신뢰도 관리부(250)는 각 작업 노드(WN)에 대한 신뢰도 정보를 획득하기 위해 글로벌 신뢰도(Global Trust) 기법을 이용한다.

클러스터 신뢰도 관리부(250)는 로컬 신뢰 의견 수집부(251), 주관 평판 계산부(252), 행동 평판 계산부(253), 신뢰도 평판 계산부(254) 및 글로벌 평판 계산부(255)를 포함할 수 있다.

로컬 신뢰 의견 수집부(251)는 클러스터내의 다수의 작업 노드(WN) 각각이 다른 작업 노드들에 대해 계산한 로컬 신뢰 의견(LTO)을 인가받아 저장한다.

주관 평판 계산부(252)는 수집된 로컬 신뢰 의견(LTO)을 기반으로 각 작업 노드(WN)에 대한 신뢰도 평가 결과인 주관 평판(Subjective Reputation: SR)을 수학적 식 2에 따라 계산한다.

수학적 식 2

$$SR_{w,u} = \sum_{j \in S_u} LTO_{j,u} \frac{HR_j \text{sim}(w,j)}{\sum_{j \in S_u} HR_j \text{sim}(w,j)}$$

여기서 S_u 는 클러스터 내의 제u 작업 노드(WN_u)에 대한 로컬 신뢰 의견(LTO)을 전송한 작업 노드 집합을 나타내고, $\text{sim}(w,j)$ 는 제w 작업 노드(WN_w)와 제j 작업 노드(WN_j)가 보고한 로컬 신뢰 의견(LTO) 사이의 코사인 유사도

를 나타내며, HR_j 는 로컬 신뢰 의견(LTO)을 생성한 각 작업 노드의 계층별 가중치로서 일반 작업 노드(WN)는 1 이고, 클러스터 헤드 노드(CH)는 1보다 큰 기지정된 값을 갖는다.

즉 주관 평판(SR)은 특정 작업 노드(WN_w)가 제u 작업 노드(WN_u)에 대해 전송한 로컬 신뢰 의견(LTO)이 다른 작업 노드(WN_j)가 전송한 로컬 신뢰 의견(LTO)과 유사할수록 높은 가중치가 가중되어 획득된다.

행동 평판 계산부(253)는 주관 평판 계산부(252)에서 계산된 주관 평판(SR)을 기반으로 신뢰 정족수(quorum) 집합(D)을 구성한다. 행동 평판 계산부(253)는 주관 평판 계산부(252)에서 계산된 SR이 유사한 노드들을 그룹화 하여 신뢰 정족수 집합(D)을 생성할 수 있다. 이때, 생성된 신뢰 정족수 집합(D)에는 해당 클러스터에 포함된 전체 작업 노드 수의 1/2 이상의 작업 노드가 포함되어야 한다. 즉 행동 평판 계산부(253)는 해당 클러스터에서 주관 평판(SR)이 유사한 과반 이상의 작업 노드를 그룹화한다.

행동 평판 계산부(253)는 클러스터에서 신뢰도 정보를 악의적으로 조작하고자 하는 작업 노드를 제외하여 정확한 신뢰도 정보를 획득하기 위해 신뢰 정족수 집합(D)을 구성하는 것으로 볼 수 있다.

신뢰 정족수 집합(D)이 구성되면, 행동 평판 계산부(253)는 신뢰 정족수 집합(D)에 포함된 작업 노드들의 주관 평판(SR)을 이용하여, 클러스터 내의 각 작업 노드에 대한 행동 평판(Behavior Reputation: BR)을 수학적 식 3에 따라 계산한다.

수학적 식 3

$$BR_{CH_i,u} = \frac{\sum_{w \in D} SR_{w,u}}{|D_{CH_i}|}$$

여기서 $|D_{CH_i}|$ 는 제i 클러스터(CL_i)의 클러스터 헤드(CH_i)가 구성한 신뢰 정족수 집합(D)에 포함된 작업 노드의 개수를 나타낸다.

그리고 신뢰도 평판 계산부(254)는 로컬 신뢰 의견(LTO)과 행동 평판(BR)을 기반으로 신뢰도 평판(Credibility Reputation: CR)을 수학적 식 4에 따라 계산한다.

수학적 식 4

$$CR_{CH_i,u} = 1 - \sqrt{\frac{\sum_{j \in \{LTO_{u,j} \neq null\}} (LTO_{u,j} - BR_{CH_i,u})^2}{|j \in \{LTO_{u,j} \neq null\}|}}$$

신뢰도 평판(CR)은 특정 작업 노드(WN_u)가 보고한 로컬 신뢰 의견(LTO)과 계산된 행동 평판(BR) 사이의 유사도를 측정하여 유사할수록 높은 값으로 획득된다. 즉 로컬 신뢰 의견(LTO)을 생성한 작업 노드의 신뢰도를 계산한다.

그리고 글로벌 평판 계산부(255)는 각 작업 노드(WN)가 빅데이터를 처리한 행동에 대한 행동 평판(BR)과 다른 작업 노드를 평가한 행동에 대한 신뢰도 평판(CR)에 따라 각 작업 노드(WN)에 대한 글로벌 평판(global reputation: GR)을 수학적 식 5에 따라 계산한다.

수학적 식 5

$$GR_{CH_i,u} = \rho BR_{CH_i,u} + (1 - \rho) CR_{CH_i,u}$$

여기서 ρ 는 평판 가중치로서 [0,1] 사이의 값을 갖는다.

즉 클러스터 신뢰도 관리부(250)는 클러스터 내의 각 작업 노드(WN)가 빅데이터를 정상적으로 처리하였는지 여

부뿐만 아니라 다른 작업 노드(WN)를 정직하게 평가하였는지 여부로 해당 작업 노드의 글로벌 평판(GR)을 결정한다.

[0083] 그리고 클러스터 내의 각 작업 노드에 대해 계산된 글로벌 평판(GR)은 해당 작업 노드의 신뢰도 정보로서, 마스터 노드(MN)로 보고된다.

[0084] 도 4는 도 2의 클러스터 설정부의 상세 구성의 일 예를 나타내고, 도 5는 작업 노드 개수와 클러스터 개수에 따른 계산 복잡도 관계를 나타내는 그래프이며, 도 6은 마스터 노드가 다수의 작업 노드 중 클러스터 헤드 노드를 지정하기 위해 전송하는 메시지의 일 예를 나타낸다.

[0085] 도 4를 참조하면, 클러스터 설정부(120)는 클러스터 재구성 판별부(121), 최적 클러스터 결정부(122), 작업 노드 클러스터링부(123) 및 클러스터 헤드 선택부(124)를 포함할 수 있다.

[0086] 클러스터 재구성 판별부(121)는 클러스터 재구성 여부를 판별한다. 클러스터 재구성 판별부(121)는 상기한 바와 같이, 마스터 노드(MN)의 초기 구동 시와 기지정된 주기에 따라 클러스터 재구성 여부를 판별할 수 있다. 또한 클러스터 재구성 판별부(121)는 빅데이터 시스템에 포함된 작업 노드(MN)의 추가 및 제거 또는 악의적 작업 노드(WN^M)으로 변경된 경우와 같은 빅데이터 시스템의 환경 요인에 변화가 발생되었는지 확인하여, 환경 변화가 발생된 것으로 확인되면 클러스터가 재구성되도록 한다.

[0087] 최적 클러스터 결정부(122)는 빅데이터 시스템에 포함된 작업 노드(WN)의 전체 개수(n)를 기반으로, n 개의 작업 노드(WN)를 다수의 클러스터(CL₁ ~ CL_k)로 클러스터링하는 경우, 각 작업 노드에 대한 신뢰도 정보를 획득하기 위한 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수(k_{opt})를 계산한다.

[0088] 상기한 바와 같이, 본 실시예에서 각 클러스터(CL₁ ~ CL_k)에 포함되는 작업 노드(WN)들은 동일한 클러스터 내에 포함되는 다른 작업 노드들에 대한 로컬 신뢰 의견(LTO)을 계산하여 클러스터 헤드 노드(CH)로 전달하고, 클러스터 헤드 노드(CH)는 다수의 작업 노드에서 전달된 로컬 신뢰 의견(LTO)을 기반으로 각 작업 노드(WN)에 대한 신뢰도 정보인 글로벌 평판(GR)을 계산한다. 그리고 또한 마스터 노드(MN)는 다수의 클러스터 헤드 노드(CH)가 보고한 각 작업 노드들에 대한 신뢰도 정보를 다시 검증하는 작업을 수행한다. 따라서 각 작업 노드들에 대한 신뢰도 정보를 생성하고 검증하는 과정에서 대량의 연산이 수행된다. 특히 빅데이터 시스템에 포함되는 작업 노드(WN)의 개수가 증가할수록, 빅데이터는 분산 처리되어 처리 시간이 감소되는 반면, 신뢰도 계산은 작업 노드(WN)의 개수에 따라 증가되므로 빅데이터 처리의 지연 요소가 될 수 있다.

[0089] 따라서 최적 클러스터 결정부(122)는 작업 노드(WN)의 전체 개수(n)에 따른 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수(k_{opt})를 계산하여 신뢰도 계산에 따른 빅데이터 처리 지연이 최대한 억제되도록 한다.

[0090] 이를 위해 우선 신뢰도 계산 복잡도를 살펴본다.

[0091] 빅데이터 시스템의 전체 n개의 작업 노드(WN₁ ~ WN_n)들을 k개의 클러스터(CL₁ ~ CL_k)로 균등하게 구분하는 경우를 가정하면, 각 클러스터(CL₁ ~ CL_k)에는 최대 $\lceil \frac{n}{k} \rceil$ 개의 작업 노드(WN)가 포함된다. 여기서 $\lceil x \rceil$ 는 보다 크거나 같은 정수 중 가장 작은 정수를 나타내는 천장함수(ceiling function)이다.

[0092] 우선 클러스터(CL₁ ~ CL_k) 내부에서는 각 작업 노드가 로컬 신뢰 의견(OLT)을 생성하므로 계산 복잡도는

$O\left(\left\lceil \frac{n}{k} \right\rceil^2\right)$ 이다. 그리고 각 작업 노드(WN)에서 생성된 로컬 신뢰 의견(OLT)을 기반으로 신뢰 정족수 집합(D)을 구성하여 로컬 신뢰 의견(OLT)의 정확도를 판단하므로 각 클러스터 내부의 신뢰도 계산 복잡도는 $O\left(\left\lceil \frac{n}{k} \right\rceil^2 \times \left\lceil \frac{n}{k} \right\rceil\right) = O\left(\left\lceil \frac{n}{k} \right\rceil^3\right)$ 이다. 한편 마스터 노드(MN)는 다수의 클러스터 헤드 노드(CH)로부터 보고된 신뢰도 정보에 대한 정확성을 판단하는 검증 작업을 클러스터 내부에서 수행된 방식과 유사하게 한번더 수행하게 되며 이때의 계산 복잡도는 $O(k^3)$ 이다. 따라서 빅데이터 시스템의 전체 신뢰도 계산 복잡도는 $O\left(\left\lceil \frac{n}{k} \right\rceil^3 + k^3\right)$ 이다.

[0093] 이에 신뢰도 계산 복잡도를 최소화하는 값(k')은 수학적 식 6에 따라 계산될 수 있다.

수학식 6

$$k' = \arg_k \left[\frac{\partial O \left(\left[\frac{n}{k} \right]^3 + k^3 \right)}{\partial k} = 0 \right] = \arg_k \left[-\frac{3n^3}{k^4} + 3k^2 = 0 \right]$$

[0094]

[0095] 수학식 6을 만족하는 k' 은 $\pm\sqrt{n}$ 이고, n 과 k 는 각각 0보다 커야 하므로($n > 0$, $k > 0$), 계산 복잡도를 최소화 하는 값(k')은 $k' = \sqrt{n}$ 이다.

[0096] 한편, 계산 복잡도($O \left(\left[\frac{n}{k} \right]^3 + k^3 \right)$)에 대한 이계 도함수는 수학식 7과 같다.

수학식 7

$$\frac{\partial^2}{\partial k^2} O \left(\left[\frac{n}{k} \right]^3 + k^3 \right) = \frac{12n^3}{k^5} + 6k$$

[0097]

[0098] 여기서도 $n > 0$, $k > 0$ 이고, 계산 복잡도 또한 $O \left(\left[\frac{n}{k} \right]^3 + k^3 \right) > 0$ 이므로, 수학식 7은 컨벡스 함수(convex function)이다.

[0099] 또한 k 는 클러스터 개수를 의미하므로 정수이어야 한다. 따라서 최적 클러스터 결정부(122)는 계산 복잡도를 최소화하는 최적 클러스터 개수(k_{opt})를 $nint(\sqrt{n})$ 으로 계산한다. 여기서 $nint(x)$ 는 x 에 가장 가까운 정수를 나타내는 함수를 의미한다.

[0100] 도 5에 도시된 빅데이터 시스템의 전체 작업 노드 개수(n)와 클러스터 개수(k)에 따른 계산 복잡도 관계를 나타 내는 그래프를 살펴보면, 클러스터의 개수(k)가 1부터 점차 증가할수록 계산 복잡도(C_{total})는 점차 감소하다가 일정 지점으로부터 다시 증가하는 형태로 나타난다. 이때 계산 복잡도가 최소인 위치가 빅데이터 시스템의 최적 클러스터 개수(k_{opt})이다. 전체 작업 노드 개수(n)에 따른 최적 클러스터 개수(k_{opt})를 연결하면 도 5의 점선 으로 나타나게 된다.

[0101] 작업 노드 클러스터링부(123)는 최적 클러스터 결정부(122)에서 계산된 최적 클러스터 개수(k_{opt})에 따라 n 개의 작업 노드를 클러스터링하여 최적 클러스터 개수(k_{opt})의 클러스터로 구분한다. 이때 작업 노드 클러스터링부 (123)는 각 클러스터에 가능한 균등한 개수의 작업 노드(WN)가 포함되도록 클러스터링을 수행한다.

[0102] 클러스터 헤드 선택부(124)는 구분된 최적 클러스터 개수(k_{opt})의 클러스터 각각에서 신뢰도 정보에 따라 하나의 작업 노드를 클러스터 헤드 노드(CH)를 선택하고, 선택된 작업 노드로 클러스터 헤드 노드(CH)로 선택되었음을 통지한다.

[0103] 클러스터 헤드 선택부(124)는 신뢰도 관리 데이터베이스(DB)에 저장된 신뢰도 정보에 따라 각 클러스터에서 다 수의 작업 노드(WN) 중 신뢰도, 즉 글로벌 평판(GR)이 가장 높은 작업 노드를 판별하고, 판별된 작업 노드를 클 러스터 헤드 노드(CH)로 선택할 수 있다.

[0104] 그리고 클러스터 헤드 선택부(124)는 선택된 작업 노드로 메시지를 전송하여, 선택된 작업 노드가 클러스터 헤 드 노드(CH)로 선택되었음을 인지할 수 있도록 한다.

[0105] 도 6을 참조하면, 마스터 노드(MN)는 클러스터를 재구성하는 경우, IP 패킷(300)의 플래그 필드(310)에서 사용 하지 않는 필드를 클러스터(Clu) 필드(311)로 이용한다. 이때 플래그 필드(310)의 나머지 DF(Don't Fragment)

필드와 MF(More Fragment) 필드는 기존의 용도 그대로 이용한다. 마스터 노드(MN)는 일반 작업 노드와의 통신 시에는 아래 Flags 이미지의 상단에 나타난 바와 같이, 플래그 필드(310)에서 클러스터 필드(311)를 0으로 설정하여 패킷을 전송하는 반면, 클러스터 헤드 노드(CH)와 통신 시에는 클러스터 필드(311)를 1로 설정하여 패킷을 전송할 수 있다.

[0106] 이와 함께 마스터 노드(MN)는 아래 Option 이미지에 도시된 바와 같이 옵션 필드(320)에 패킷을 수신하는 작업 노드의 클러스터 소속 정보, 클러스터 헤드 노드 지정 여부, 클러스터를 관리하기 위해 새로 지정된 클러스터 헤드 노드(CH) 정보 등을 포함시켜 전송함으로써, 각 클러스터에서 각 작업 노드(WN)가 클러스터 헤드 노드(CH)로 선택되었는지 여부와 자신이 포함된 클러스터의 클러스터 헤드 노드(CH)를 인지할 수 있도록 한다.

[0107] 여기서 옵션 필드(320)는 일 예로 96 비트로 구성될 수 있다. 옵션 필드(320)에서 첫 16비트는 마스터 노드(MN)가 작업 노드(WN)들이 속한 클러스터(CL)를 구분하기 위한 클러스터 정보로 구성될 수 있다. 그리고 두번째 16비트는 클러스터(CL)를 관리하는 클러스터 헤드 노드(CH)를 위한 정보로 구성될 수 있다. 여기서는 패킷을 수신하는 작업 노드(WN)가 클러스터 헤드 노드(CH)로 지정된다면, 해당 필드의 첫 비트는 1로 변경되어, 해당 노드가 클러스터 헤드 노드(CH)로 지정되었음을 인지할 수 있도록 한다. 이후 15비트는 해당 클러스터에 속한 작업 노드(WN)의 수를 표현한다. 다음 32비트는 클러스터 헤드 노드(CH)로 지정되지 않은 나머지 작업 노드(WN)들 위해, 자신이 속한 클러스터의 클러스터 헤드 노드(CH)에 대한 정보가 포함될 수 있다. 이는 작업 노드(WN)들이 클러스터 관리 노드에 연결될 수 있도록 하기 위한 정보로서, 클러스터 헤드 노드(CH)의 IP 주소 등과 같은 정보가 포함될 수 있다. 다음 32비트는 동일한 클러스터에 포함된 클러스터 헤드 노드(CH)와 작업 노드(WN) 사이의 인증을 위한 정보가 포함될 수 있다. 이 정보들은 시스템에서 미리 지정되어, 작업 노드(WN)들이 다른 클러스터에 연결되는 것을 방지하도록 한다.

[0108] 도 6에 도시된 메시지는 일 예로서 빅데이터 시스템에 따라 메시지 형태는 변경될 수 있다.

[0109] 도 7은 도 2의 신뢰도 기반 판별부의 상세 구성의 일 예를 나타낸다.

[0110] 도 7을 참조하면 신뢰도 기반 판별부는 신뢰도 정보 검증부(131), 악의적 노드 선별부(132) 및 신뢰 노드수 확인부(133)를 포함할 수 있다.

[0111] 신뢰도 정보 검증부(131)는 최적 클러스터 개수(k_{opt})에 따라 클러스터링된 각 클러스터의 클러스터 헤드 노드(CH)로부터 보고된 각 작업 노드(WN)에 대한 신뢰도 정보, 즉 글로벌 평판(GR)에 대한 검증을 미리 지정된 수행한다. 여기서 신뢰도 정보 검증부(131)는 클러스터 헤드 노드(CH)가 각 작업 노드(WN)에 대해 수행한 신뢰도 검증과 동일한 방식으로 각 작업 노드(WN)에 대한 글로벌 평판(GR)을 재계산하여 비교함으로써 신뢰도 정보를 검증할 수 있다.

[0112] 신뢰도 정보 검증부(131)는 클러스터 헤드 노드(CH)에서 보고된 신뢰도 정보를 검증한 결과, 신뢰도 정보가 신뢰할 수 있는 수준인 것으로 판별되면, 보고된 신뢰도 정보를 신뢰도 관리 데이터베이스(DB)로 전송하여 저장한다.

[0113] 악의적 노드 선별부(132)는 검증된 각 작업 노드(WN)에 대한 신뢰도 정보를 확인하여, 신뢰도 정보, 즉 글로벌 평판(GR)이 기지정된 기준값 미만인 작업 노드들을 악의적 작업 노드(WN^M)로 선별한다. 그리고 나머지 작업 노드들, 즉 글로벌 평판(GR)이 기준값 이상인 작업 노드들을 정직 작업 노드(WN^H)로 구분한다.

[0114] 신뢰 노드수 확인부(133)는 빅데이터 시스템의 전체 작업 노드 개수(n) 중에서 악의적 노드 선별부(132)에서 구분된 악의적 작업 노드(WN^M)의 개수(N^M)와 정직 작업 노드(WN^H)의 개수(N^H)를 판별한다.

[0115] 도 8은 도 2의 최적 개수 판별부의 상세 구성의 일 예를 나타낸다.

[0116] 도 8을 참조하면, 최적 개수 판별부(140)는 전송 시간 계산부(141), 작업 시간 계산부(142) 및 최적 개수 계산부(143)를 포함할 수 있다.

[0117] 전송 시간 계산부(141)는 마스터 노드(MN)에서 빅데이터를 처리하기 위해 소요되는 마스터 처리 시간(T_m)을 계산한다.

[0118] 상기한 바와 같이, 마스터 노드(MN)는 실시간으로 전송된 빅데이터를 최적 클러스터 개수(k_{opt})에 따라 구분된 다수의 클러스터(CL)의 클러스터 헤드 노드(CH)로 분배하여 전송한다. 이때, 마스터 처리 시간(T_m)은 수학적 8

과 같다.

수학식 8

$$T_m = T_{vs}^m + T_{comm}^m + T_{add}^m = kiA + \frac{Bs}{k} + P_{e_{comm}} \frac{Bs}{k}$$

[0119]

[0120] 여기서 T_{vs}^m 는 마스터 노드(MN)와 클러스터 헤드 노드(CH) 사이의 변수 공유(variable sharing) 시간을 나타내고, T_{comm}^m 은 마스터 노드(MN)가 클러스터 헤드 노드(CH)로 데이터를 전송하는데 소요되는 데이터 전송 시간을 나타낸다. 그리고 T_{add}^m 는 마스터 노드(MN)가 클러스터 헤드 노드(CH)로 데이터를 전송할 때 발생하는 전송 오류를 복구하기 위해 데이터를 재전송하는데 소요되는 오류 복구 시간이다. 즉 마스터 처리 시간(T_m)은 마스터 노드(MN)와 클러스터 헤드 노드(CH) 사이의 변수 공유 시간(T_{vs}^m)과 데이터 전송 시간(T_{comm}^m) 및 오류 복구 시간(T_{add}^m)의 합으로 계산될 수 있다.

[0121]

한편, $A = \theta_{vs} T_{vs}^{baseline}$ 이고, $B = \theta_{comm} T_{comm}^{baseline}$ 로서, $T_{vs}^{baseline}$ 과 $T_{comm}^{baseline}$ 는 각각 단위 크기의 변수 및 단위 크기의 데이터를 전송하는데 소요되는 단위 전송 시간이며, θ_{vs} 와 θ_{comm} 은 각각 기지정된 계수이다. 그리고 $P_{e_{comm}}$ 는 전송 오류 발생 확률이다. 그리고 k 는 클러스터 개수이고, i 는 반복 횟수이며, s 는 데이터 크기를 나타낸다.

[0122]

그리고 작업 시간 계산부(142)는 각 클러스터($CL_1 \sim CL_k$)에서 분배된 빅데이터를 처리하기 위해 소요되는 클러스터 처리 시간(T_c^k)을 계산한다.

[0123]

클러스터 처리 시간(T_c^k)은 마스터 처리 시간(T_m)과 유사하게 수학식 9로 계산될 수 있다.

수학식 9

$$T_c^k = T_{vs}^k + T_{comm}^k + T_{exec}^k + T_{add}^k$$

[0124]

[0125] 여기서 T_{vs}^k 는 클러스터 헤드 노드(CH)와 작업 노드(WN) 사이의 변수 공유 시간을 나타내고, T_{comm}^k 는 클러스터 헤드 노드(CH)와 작업 노드(WN) 사이의 데이터 전송 시간을 나타내며, T_{exec}^k 는 작업 노드(WN)가 인가된 데이터를 처리하는 데이터 처리 시간을 나타낸다. 그리고 T_{add}^k 는 오류 복구 시간으로, 통신 오류로 인한 통신 오류 복구 시간과 함께 작업 노드(WN)가 데이터를 처리하는 과정에서 발생하는 오류를 복구하기 위한 데이터 처리 복구 시간이 포함될 수 있다.

[0126]

따라서 클러스터 처리 시간(T_c^k)은 변수 공유 시간(T_{vs}^k)과 데이터 전송 시간(T_{comm}^k), 오류 복구 시간(T_{add}^k) 및 데이터 처리 시간(T_{exec}^k)의 합으로 계산될 수 있다.

[0127]

그리고 수학식 9의 클러스터 처리 시간(T_c^k)은 제 k 클러스터(CL_k)에 포함된 작업 노드의 개수(n_k)에 따른 제 k 클러스터(CL_k)의 클러스터 처리 시간(T_c^k)은 수학식 10과 같이 다시 표현될 수 있다.

수학식 10

$$\begin{aligned}
 T_c^k &= T_{vs}^k + T_{comm}^k + T_{exec}^k + T_{add}^k \\
 &= n_k iA + \frac{Bs \frac{n_k}{n}}{n_k} + \frac{iM_a \frac{n_k}{n}}{n_k} + P_{e_{comm}} \frac{Bs \frac{n_k}{n}}{n_k} + P_{e_{exec}} \left(n_k iA + \frac{Bs \frac{n_k}{n}}{n_k n_k} + \frac{iM_a \frac{n_k}{n}}{n_k n_k} \right) \\
 &= n_k iA + \frac{Bs}{n} + \frac{iM_a}{n} + P_{e_{comm}} \frac{Bs}{n} + P_{e_{exec}} \left(n_k iA + \frac{Bs}{nn_k} + \frac{iM_a}{nn_k} \right)
 \end{aligned}$$

[0128]

[0129] 여기서 Ma 는 작업 노드(WN)의 단위 크기 데이터 처리 시간을 나타내고, $P_{e_{exec}}$ 는 데이터 처리 오류율을 나타낸다.

[0130] 최적 개수 계산부(143)는 전송 시간 계산부(141)에서 계산된 마스터 처리 시간(T_m)과 작업 시간 계산부(142)에서 계산된 클러스터 처리 시간(T_c^k)의 합으로 계산되는 빅데이터 처리 시간(T_p)이 작업 명령에서 요청된 목표 시간(T_{object})을 만족시킬 수 있도록 하는 최소의 작업 노드 개수를 최적 작업 노드 개수(N_{opt})로 계산한다.

[0131] 최적 개수 계산부(143)는 마스터 처리 시간(T_m)과 클러스터 처리 시간(T_c^k)의 합으로 계산되는 빅데이터 처리 시간($T_p = T_m + T_c^k$)은 수학식 11에 따라 계산될 수 있다.

수학식 11

$$\begin{aligned}
 T_p &= T_m + T_c^k \\
 &= (T_{vs}^m + T_{comm}^m + T_{add}^m) + (T_{vs}^k + T_{comm}^k + T_{exec}^k + T_{add}^k) \\
 &= k iA + \frac{Bs}{k} + P_{e_{comm}} \frac{Bs}{k} + n_k iA + \frac{Bs}{n} + \frac{iM_a}{n} \\
 &\quad + P_{e_{comm}} \frac{Bs}{n} + P_{e_{exec}} \left(n_k iA + \frac{Bs}{nn_k} + \frac{iM_a}{nn_k} \right)
 \end{aligned}$$

[0132]

[0133] 그리고 빅데이터 처리 시간(T_p)은 작업 명령에서 요청된 목표 시간(T_{object}) 이하($T_p \leq T_{object}$)이어야 하며, 이에 수학식 11은 수학식 12로 다시 표현될 수 있다.

수학식 12

$$\begin{aligned}
 T_p &= (T_{vs}^m + T_{comm}^m + T_{add}^m) + (T_{vs}^k + T_{comm}^k + T_{exec}^k + T_{add}^k) \\
 &= k iA + \frac{Bs}{k} + P_{e_{comm}} \frac{Bs}{k} + n_k iA + \frac{Bs}{n} + \frac{iM_a}{n} + P_{e_{comm}} \frac{Bs}{n} \\
 &\quad + P_{e_{exec}} \left(n_k iA + \frac{Bs}{nn_k} + \frac{iM_a}{nn_k} \right) \leq T_{object}
 \end{aligned}$$

[0134]

[0135] 그리고 수학식 12를 작업 노드의 개수(n)에 대해 정리하면, 수학식 12는 수학식 13으로 획득될 수 있다.

수학식 13

$$\begin{aligned} & n \left(\frac{1}{k} (1 + P_{e_{\text{exec}}}) iA \right) + \left(kiA + \frac{Bs}{k} + P_{e_{\text{comm}}} \frac{Bs}{k} - T_{\text{object}} \right) \\ & + \frac{1}{n} (Bs + iM_a + P_{e_{\text{comm}}} Bs) \\ & + \frac{1}{n^2} (kP_{e_{\text{exec}}} (Bs + iM_a)) \leq 0 \end{aligned}$$

[0136]

[0137] 작업 노드의 개수(n)는 $n \geq 0$ 이어야 하므로, 수학식 13은 수학식 14로 다시 획득될 수 있다.

수학식 14

$$\begin{aligned} & n^3 \left(\frac{1}{k} (1 + P_{e_{\text{exec}}}) iA \right) + n^2 \left(kiA + \frac{Bs}{k} + P_{e_{\text{comm}}} \frac{Bs}{k} - T_{\text{object}} \right) \\ & + n (Bs + iM_a + P_{e_{\text{comm}}} Bs) + kP_{e_{\text{exec}}} (Bs + iM_a) \\ & = xn^3 + yn^2 + zn + w \leq 0 \end{aligned}$$

[0138]

[0139] 여기서 $x = \frac{1}{k} (1 + P_{e_{\text{exec}}}) iA$, $y = kiA + \frac{Bs}{k} + P_{e_{\text{comm}}} \frac{Bs}{k} - T_{\text{object}}$, $z = Bs + iM_a + P_{e_{\text{comm}}} Bs$ 및 $w = kP_{e_{\text{exec}}} (Bs + iM_a)$ 이다.

[0140] 수학식 14의 3차 방정식의 해를 n_1 , n_2 , n_3 라고 할 때, n_1 , n_2 , n_3 는 각각 수학식 15 내지 17로 계산된다.

수학식 15

$$n_1 = -\frac{y}{3x} - \frac{1}{3x} \left[\frac{\Psi + \{\Psi^2 - 4(y^2 - 3xz)^3\}^{\frac{1}{2}}}{2} \right]^{\frac{1}{3}} - \frac{1}{3x} \left[\frac{\Psi - \{\Psi^2 - 4(y^2 - 3xz)^3\}^{\frac{1}{2}}}{2} \right]^{\frac{1}{3}}$$

[0141]

수학식 16

$$n_2 = -\frac{y}{3x} + \frac{1 + i\sqrt{3}}{6x} \left[\frac{\Psi + \{\Psi^2 - 4(y^2 - 3xz)^3\}^{\frac{1}{2}}}{2} \right]^{\frac{1}{3}} + \frac{1 - i\sqrt{3}}{6x} \left[\frac{\Psi - \{\Psi^2 - 4(y^2 - 3xz)^3\}^{\frac{1}{2}}}{2} \right]^{\frac{1}{3}}$$

[0142]

수학식 17

$$n_3 = -\frac{y}{3x} + \frac{1 - i\sqrt{3}}{6x} \left[\frac{\Psi + \{\Psi^2 - 4(y^2 - 3xz)^3\}^{\frac{1}{2}}}{2} \right]^{\frac{1}{3}} + \frac{1 + i\sqrt{3}}{6x} \left[\frac{\Psi - \{\Psi^2 - 4(y^2 - 3xz)^3\}^{\frac{1}{2}}}{2} \right]^{\frac{1}{3}}$$

[0143]

[0144] 수학식 15 내지 17에서 $\Psi = 2y^3 - 9xyz + 27x^2w$ 이다.

[0145] 최적 개수 계산부(143)는 수학식 15 내지 17에 따라 계산되는 3차 방정식의 3개의 해(n_1 , n_2 , n_3)에서 수학식 18

에 따라 최소 실수를 탐색하고, 탐색된 최소 실수 이상인 가장 작은 정수를 최적 작업 노드 개수(N_{opt})로서 획득한다.

수학식 18

$$N_{opt} = \{ \lceil \min(n_i) \rceil \mid n_i > 0, n_i \in \mathbb{R}, i = 1, 2, 3 \}$$

도 2에 도시된 작업 노드수 결정부(150)는 신뢰도 기반 판별부(130)에서 판별된 정직 작업 노드(WN^H)의 개수(N^H)와 악의적 작업 노드(WN^M)의 개수(N^M)와 최적 개수 판별부(140)에서 판별된 최적 작업 노드 개수(N_{opt})를 기반으로 처리 노드 수(N_{pro})를 결정한다.

여기서 작업 노드수 결정부(150)는 정직 작업 노드(WN^H)의 개수(N^H)가 최적 작업 노드 개수(N_{opt}) 이상이면, 최적 작업 노드 개수(N_{opt})를 처리 노드 수(N_{pro})로 결정($N_{pro} = N_{opt}$)한다. 그러나 정직 작업 노드(WN^H)의 개수(N^H)가 최적 작업 노드 개수(N_{opt})보다 작으면, 정직 작업 노드(WN^H)의 개수(N^H)를 처리 노드 수(N_{pro})로 결정($N_{pro} = N^H$)한다.

도 9는 본 발명의 일 실시예에 따른 빅데이터 처리 제어 방법을 나타낸다.

본 실시예의 빅데이터 처리 방법은 도 1에 도시된 빅데이터 시스템에서 빅데이터 처리 제어 장치로 동작하여, 빅데이터 시스템에 포함된 n 개의 작업 노드(WN)를 제어하는 마스터 노드(MN)에 의해 수행된다.

도 9를 참조하면, 도 9의 빅데이터 처리 방법은 우선 다수의 센서 또는 단말에서 전송된 데이터를 수집하여 빅데이터를 구성한다(S11). 그리고 사용자 단말(User)로부터 빅데이터를 이용한 작업 명령이 인가되는지 판별한다(S20). 만일 작업 명령이 인가되면, 클러스터 재구성 단계(S30)를 수행한다. 클러스터 재구성 과정은 우선 클러스터를 재구성 해야하는지 여부를 판별한다(S31). 여기서 클러스터 재구성은 기지정된 주기로 수행될 수도 있으나, 빅데이터 시스템의 환경 변화 발생 여부에 따라 수행될 수 있다. 예로서 빅데이터 시스템에 작업 노드(WN)가 추가되거나 제거된 경우 및 정직 작업 노드(WN^H)가 악의적 작업 노드(WN^M)로 변경된 경우 또는 작업 명령에서 요청되는 목표 시간(T_{object})의 변화 등에 따라 클러스터가 재구성될 수 있다.

만일 클러스터를 재구성해야 하는 것으로 판별되면, 전체 n 개의 작업 노드(WN)를 k 개의 클러스터로 구분하는 경우, 구분된 클러스터 내부에서 수행되는 신뢰도 계산과 마스터 노드(MN)에서 수행되는 신뢰도 검증을 포함한 신뢰도 계산 복잡도가 최소가 되도록 하는 최적 클러스터 개수(k_{opt})를 계산한다(S32). 여기서 최적 클러스터 개수(k_{opt})는 $\text{nint}(\sqrt{n})$ 으로 계산될 수 있다.

그리고 계산된 최적 클러스터 개수(k_{opt})에 따라 n 개의 작업 노드(WN)를 클러스터링하여 k_{opt} 개의 클러스터(CL)로 구분한다(S33). n 개의 작업 노드(WN)가 k_{opt} 개의 클러스터(CL)로 구분되면, k_{opt} 개의 클러스터(CL) 각각에서 하나의 작업 노드를 클러스터 헤드 노드(CH)로 지정한다(S34). 이때 마스터 노드(MN)는 이전 검증되어 저장된 각 작업 노드의 신뢰도 정보에 기반하여, 각 클러스터에서 가장 높은 신뢰도를 갖는 작업 노드를 클러스터 헤드 노드(CH)로 지정할 수 있다.

한편, 클러스터가 재구성되거나, 클러스터를 재구성할 필요가 없으면, 작업 노드 신뢰도 판별 단계(S40)와 최적 빅데이터 처리 작업 노드 개수 판별 단계(S50)를 병렬로 함께 수행한다.

작업 노드 신뢰도 판별 단계(S40)에서는 우선 현재 구성된 클러스터 각각의 클러스터 헤드 노드(CH)가 대응하는 클러스터에 포함된 각 작업 노드에 대해 획득한 신뢰도 정보를 보고받아 수집한다(S41).

여기서 신뢰도 정보는 각 클러스터의 클러스터 헤드 노드(CH)가 해당 클러스터(CL) 내의 작업 노드(WN)가 다른 작업 노드의 빅데이터 처리 행동을 분석하여 수집한 로컬 신뢰 의견($LT0$)을 기반으로 획득한다.

각 클러스터 내의 다수의 작업 노드(WN) 각각은 다른 작업 노드(WN)의 빅데이터 처리 행동을 분석하여 로컬 신

뢰 의견(LTO)을 획득하고, 클러스터 헤드 노드(CH)는 각 작업 노드가 획득한 로컬 신뢰 의견(LTO)을 수집하여 글로벌 신뢰도 기법에 따라 각 작업 노드에 대한 신뢰도 정보를 획득한다. 클러스터 헤드 노드(CH)는 각 작업 노드(WN)에서 수집된 로컬 신뢰 의견(LTO) 중 동일한 작업 노드를 평가한 로컬 신뢰 의견(LTO) 사이의 유사도에 따라 주관 평판(SR)을 계산하고, 계산된 주관 평판(SR) 중 높은 주관 평판(SR)을 갖는 과반 이상의 작업 노드(WN)에 대한 신뢰 정족수 집합(D)을 획득한다. 또한 클러스터 헤드 노드(CH)는 신뢰 정족수 집합(D) 내의 작업 노드(WN)들의 주관 평판(SR)을 기반으로 각 작업 노드(WN)가 다른 작업 노드를 평가한 행동 평판(BR)을 계산하고, 로컬 신뢰 의견(LTO)과 행동 평판(BR)을 기반으로 로컬 신뢰 의견(LTO)을 생성한 작업 노드의 신뢰도를 나타내는 신뢰도 평판(CR)을 계산한다. 이후, 클러스터 헤드 노드(CH)는 각 작업 노드(WN)가 빅데이터를 처리한 행동에 대한 행동 평판(BR)과 다른 작업 노드를 평가한 행동에 대한 신뢰도 평판(CR)에 따라 각 작업 노드(WN)에 대한 글로벌 평판(GR)을 계산하여 신뢰도 정보로서 마스터 노드(MN)으로 보고한다. 이에 마스터 노드(MN)는 클러스터 헤드 노드(CH)가 각 작업 노드에 대해 획득한 신뢰도 정보를 인가받아 수집한다.

[0158] 마스터 노드(MN)는 각 클러스터(CL)의 클러스터 헤드 노드(CH)로부터 신뢰도 정보가 수집되면, 수집된 신뢰도 정보를 기지정된 방식에 따라 검증한다(S42). 이때 마스터 노드(MN)는 일 예로 클러스터 헤드 노드(CH)와 마찬가지로 글로벌 신뢰도 기법에 따라 각 작업 노드에 대한 신뢰도 정보를 획득하여 비교함으로써 검증을 수행할 수 있다.

[0159] 그리고 검증 결과로 클러스터 헤드 노드(CH)가 보고한 작업 노드(WN)에 대한 신뢰도 정보가 신뢰할 수 있는 것으로 판별되면, 각 작업 노드(WN)에 대한 신뢰도 정보, 즉 글로벌 평판(GR)을 기반으로 전체 n 개의 작업 노드(WN)에서 정직 작업 노드(WN^H)와 악의적 작업 노드(WN^M)를 판별하고, 정직 작업 노드(WN^H)의 개수(N^H)와 악의적 작업 노드(WN^M)의 개수(N^M)를 각각 확인한다.

[0160] 이와 함께 마스터 노드(MN)는 신뢰할 수 있는 것으로 판별된 신뢰도 정보를 신뢰도 관리 데이터베이스(DB)로 전송하여 저장함으로써, 이후 클러스터를 재구성하게 되면 클러스터 헤드 노드(CH)를 지정하기 위해 이용할 수 있도록 할 수 있다.

[0161] 한편, 최적 빅데이터 처리 작업 노드 개수 판별 단계(S50)에서는 마스터 노드(MN)에서 빅데이터를 처리하기 위해 소요되는 마스터 처리 시간(T_m)을 마스터 노드(MN)와 클러스터 헤드 노드(CH) 사이의 변수 공유 시간(T_{vs}^m)과 데이터 전송 시간(T_{comm}^m) 및 오류 복구 시간(T_{add}^m)의 합으로 계산한다(S51).

[0162] 그리고 각 클러스터($CL_1 \sim CL_k$)가 분배된 빅데이터를 처리하기 위해 소요되는 클러스터 처리 시간(T_c^k)을 변수 공유 시간(T_{vs}^k)과 데이터 전송 시간(T_{comm}^k), 오류 복구 시간(T_{add}^k) 및 데이터 처리 시간(T_{exec}^k)의 합으로 계산한다(S52).

[0163] 이후, 마스터 처리 시간(T_m)과 클러스터 처리 시간(T_c^k)의 합으로 계산되는 빅데이터 처리 시간($T_p = T_m + T_c^k$)이 작업 명령에서 요청된 목표 시간(T_{object}) 이하($T_p \leq T_{object}$)가 되도록 하는 최적 작업 노드 개수(N_{opt})를 계산한다.

[0164] 정직 작업 노드(WN^H)의 개수(N^H)와 악의적 작업 노드(WN^M)의 개수(N^M) 및 최적 작업 노드 개수(N_{opt})가 계산되면, 빅데이터를 실제로 처리할 처리 노드 수(N_{pro})를 계산한다. 빅데이터 처리 노드 수(N_{pro})는 정직 작업 노드(WN^H)의 개수(N^H)의 개수가 최적 작업 노드 개수(N_{opt}) 이상인지 판별하여, 정직 작업 노드(WN^H)의 개수(N^H)의 개수가 최적 작업 노드 개수(N_{opt}) 이상이면, 최적 작업 노드 개수(N_{opt})를 처리 노드 수(N_{pro})로 결정($N_{pro} = N_{opt}$)할 수 있다. 그러나 정직 작업 노드(WN^H)의 개수(N^H)가 최적 작업 노드 개수(N_{opt})보다 작으면, 정직 작업 노드(WN^H)의 개수(N^H)를 처리 노드 수(N_{pro})로 결정($N_{pro} = N^H$)할 수 있다.

[0165] 빅데이터를 처리하기 위한 처리 노드 수(N_{pro})에 따라 k_{opt} 개의 클러스터(CL) 각각에서 빅데이터를 처리할 작업 노드의 개수를 결정하고, 결정된 작업 노드의 개수에 따라 빅데이터를 분배하여 각 클러스터(CL)의 클러스터 헤드 노드(CH)로 분배하여 빅데이터가 처리되도록 한다(S70). 이때 마스터 노드(MN)는 각 클러스터에 포함된 정

직 작업 노드(WN^H)의 개수(N^H)를 고려하여 각 클러스터에 빅데이터를 분배할 수 있다. 또한 정직 작업 노드(WN^H)의 개수(N^H)가 충분한 경우 각 클러스터에 균등하게 빅데이터를 분배할 수도 있다. 분배된 빅데이터를 인가받은 클러스터 헤드 노드(CH)는 해당 클러스터 내에서 악의적 작업 노드(WN^M)를 배제하고, 정직 작업 노드(WN^H)로 다시 인가된 빅데이터를 재분배하여 전송하고, 빅데이터를 분배받은 정직 작업 노드(WN^H)는 인가된 빅데이터를 처리하여 클러스터 헤드 노드(CH)로 전달한다. 이에 클러스터 헤드 노드(CH)는 처리된 빅데이터를 수집하여 마스터 노드(MN)로 전달하고, 마스터 노드(MN)는 각 클러스터 헤드 노드(CH)에서 처리되어 전달된 빅데이터를 수집하여 빅데이터 처리 결과를 사용자 단말(User)로 전달한다.

[0166] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.

[0167] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

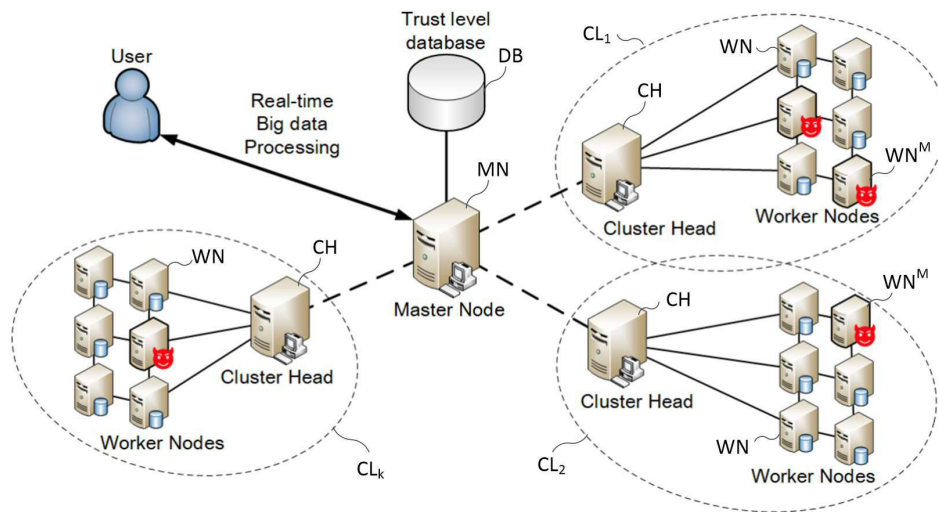
[0168] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

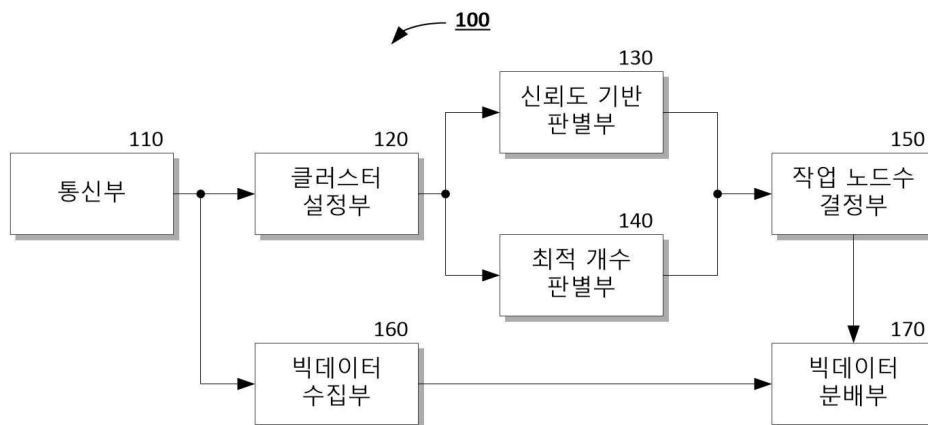
[0169] MN: 마스터 노드 DB: 신뢰도 관리 데이터베이스
 WN: 작업 노드 CL: 클러스터
 CH: 클러스터 헤드 노드 User: 사용자 단말
 110: 통신부 120: 클러스터 설정부
 121: 클러스터 재구성 판별부 122: 최적 클러스터 결정부
 123: 작업 노드 클러스터링부 124: 클러스터 헤드 선택부
 130: 신뢰도 기반 판별부 131: 신뢰도 정보 검증부
 132: 악의적 노드 선별부 133: 신뢰 노드수 확인부
 140: 최적 개수 판별부 141: 전송 시간 계산부
 142: 작업 시간 계산부 143: 최적 개수 계산부
 150: 작업 노드수 결정부 210: 작업 통신부
 220: 데이터 처리부 230: 로컬 신뢰 의견 생성부
 240: 데이터 분배부 250: 클러스터 신뢰도 관리부
 251: 로컬 신뢰 의견 수집부 252: 주관 평판 계산부
 253: 행동 평판 계산부 254: 신뢰도 평판 계산부
 255: 글로벌 평판 계산부

도면

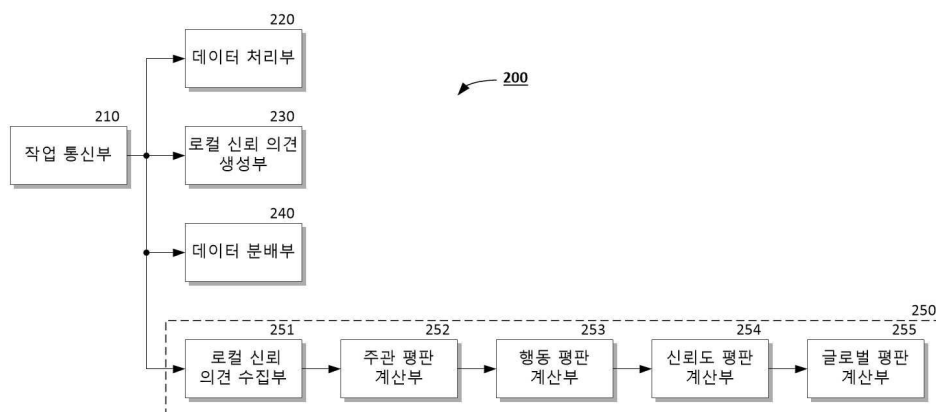
도면1



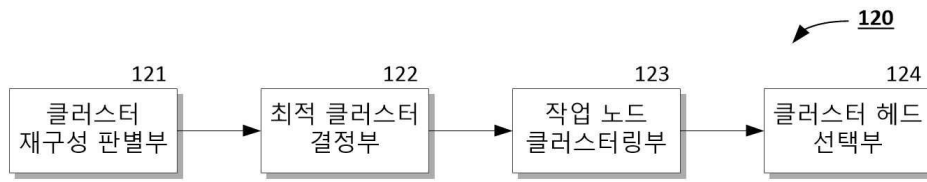
도면2



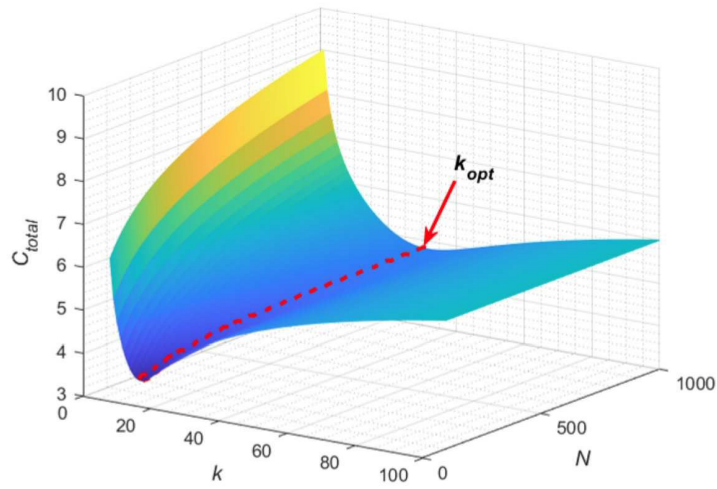
도면3



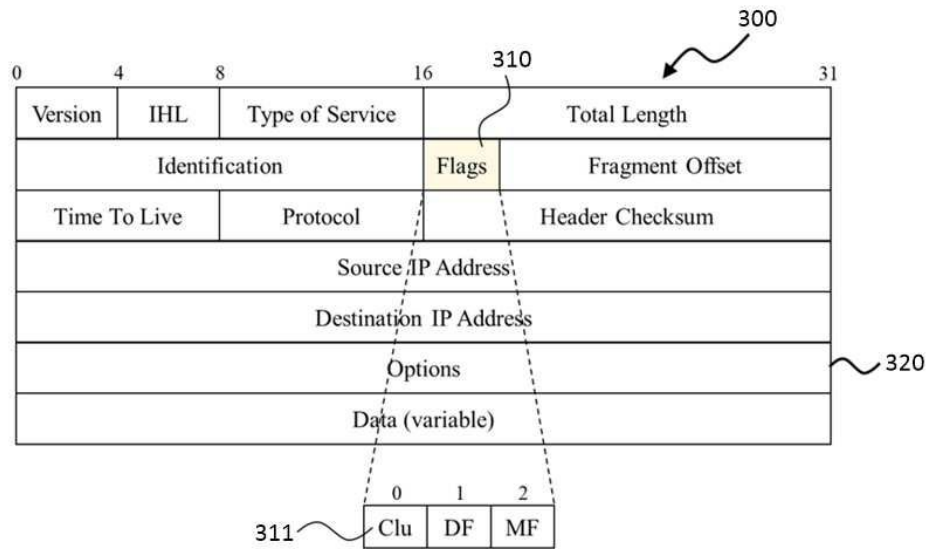
도면4



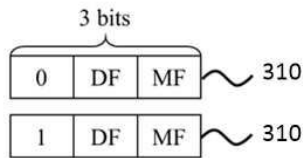
도면5



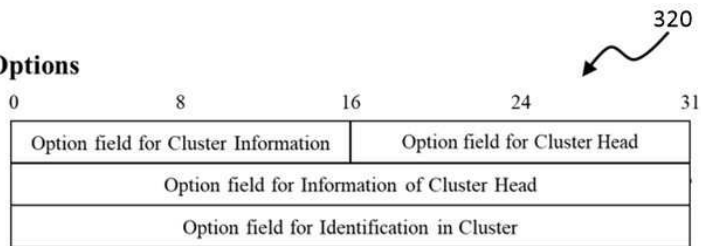
도면6



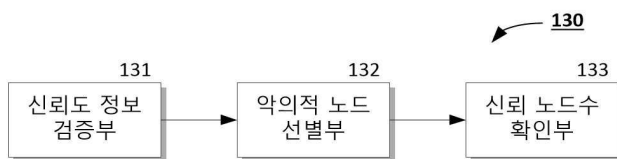
Flags



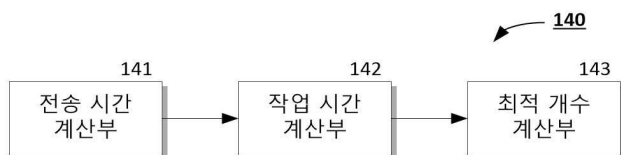
Options



도면7



도면8



도면9

