(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년12월19일
(11) 등록번호 10-1810180
(24) 등록일자 2017년12월12일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01)
(52) CPC특허분류
G06F 17/30318 (2013.01)
(21) 출원번호 10-2017-0061012
(22) 출원일자 2017년05월17일
심사청구일자 2017년05월17일
(56) 선행기술조사문헌
KR1020150030332 A*
(뒷면에 계속)

(73) 특허권자
네이버시스템(주)
서울특별시 송파구 중대로 135, 동관 16층 (가락동, 아이티벤처타워)
이치호
경상남도 김해시 진영읍 진영로 284, 103동 201호(협성휴포레 진영)
(72) 발명자
이치호
경상남도 김해시 진영읍 진영로 284, 103동 201호(협성휴포레 진영)
(74) 대리인
특허법인메이저

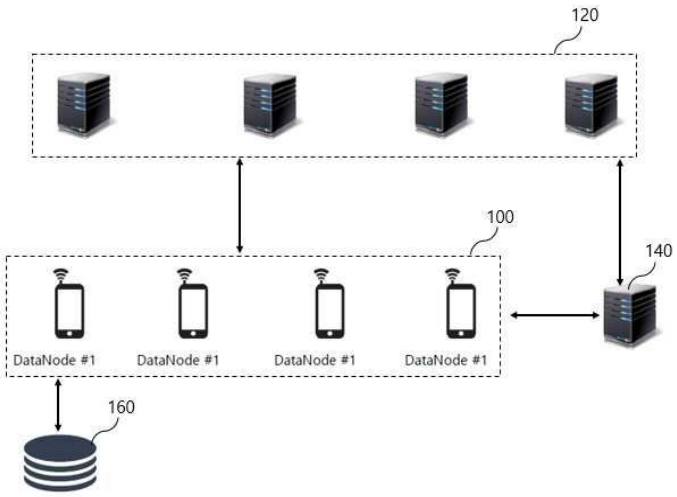
전체 청구항 수 : 총 10 항 심사관 : 최정권

(54) 발명의 명칭 **사용자 장치 기반의 빅데이터 분산 처리 방법 및 장치**

(57) 요 약

사용자 장치 기반의 빅데이터 분산 처리 방법 및 장치가 개시되어 있다. 사용자 장치 기반의 빅데이터 분산 처리 방법은 수집 서버가 수집 대상 데이터를 분할 적재하는 단계, 수집 서버는 마스터 노드에게 상기 수집 대상 데이터의 저장 상태를 전송하는 단계, 마스터 노드가 복수의 데이터 노드의 상태 정보를 기반으로 수집 대상 데이터의 처리를 복수의 데이터 노드 중 복수의 선택 데이터 노드로 요청하는 단계와 복수의 선택 데이터 노드가 수집 대상 데이터를 분산 처리하는 단계를 포함할 수 있다.

대 표 도 - 도1



(56) 선행기술조사문헌
KR1020150062596 A*
KR1020150111801 A*
US20140250153 A1
KR1020150091901 A
*는 심사관에 의하여 인용된 문헌

명 세 서

청구범위

청구항 1

사용자 장치 기반의 빅데이터 분산 처리 방법은,

수집 서버가 수집 대상 데이터를 분할 적재하는 단계;

상기 수집 서버는 마스터 노드에게 상기 수집 대상 데이터의 저장 상태를 전송하는 단계;

상기 마스터 노드가 복수의 데이터 노드의 상태 정보를 기반으로 상기 수집 대상 데이터의 처리를 상기 복수의 데이터 노드 중 복수의 선택 데이터 노드로 요청하는 단계;

상기 복수의 선택 데이터 노드가 상기 수집 대상 데이터를 분산 처리하는 단계를 포함하되,

상기 복수의 데이터 노드 각각으로서 동작하는 사용자 장치는 예측 데이터 통신량 및 예측 프로세서 사용량에 대한 정보를 고려하여 데이터 처리 가능 정보를 전송하고,

상기 예측 데이터 통신량은 상기 사용자 장치에서 구동되는 어플리케이션에서 사용할 것으로 예측되는 데이터의 크기이고,

상기 예측 프로세서 사용량은 상기 어플리케이션에서 사용할 것으로 예측되는 프로세서의 사용량이고,

상기 데이터 처리 가능 정보의 전송 주기는 상기 예측 데이터 통신량 및 상기 예측 프로세서 사용량의 변화를 기반으로 설정되는 것을 특징으로 하는 방법.

청구항 2

제1항에 있어서,

상기 복수의 데이터 노드는 상기 마스터 노드로 데이터 처리 가능 정보를 전송하는 단계를 더 포함하고,

상기 데이터 처리 가능 정보는 사용자의 상기 사용자 장치에 대한 사용 특성에 대한 정보를 더 포함하는 것을 특징으로 하는 방법.

청구항 3

제2항에 있어서,

상기 데이터 처리 가능 정보는 상기 데이터 노드의 통신 상태, CPU(central processing unit) 사용률, 메모리 사용률에 대한 정보를 포함하는 것을 특징으로 하는 방법.

청구항 4

제3항에 있어서,

상기 선택 데이터 노드는 상기 마스터 노드로부터 수신한 푸시 알림 정보에 포함된 데이터 처리 지시를 기반으로 할당 데이터를 상기 수집 서버로부터 수신하여 처리하고 처리 결과를 서비스 데이터베이스로 전송하는 것을 특징으로 하는 방법.

청구항 5

제4항에 있어서,

상기 마스터 노드는 특정 시간까지 상기 할당 데이터가 처리되지 않은 미처리 데이터 노드를 파악하고, 상기 미처리 데이터 노드로 푸시 알림으로 재요청하거나 또는 상기 미처리 데이터 노드의 임무 취소 후 다른 데이터 노드에게 해당 임무를 지시하는 것을 특징으로 하는 방법.

청구항 6

사용자 장치 기반의 빅데이터 분산 처리 시스템은,

수집 대상 데이터를 분할 적재하는 단계; 마스터 노드에게 상기 수집 대상 데이터의 저장 상태를 전송하도록 구현되는 수집 서버;

복수의 데이터 노드의 상태 정보를 기반으로 상기 수집 대상 데이터의 처리를 상기 복수의 데이터 노드 중 복수의 선택 데이터 노드로 요청하도록 구현되는 마스터 노드; 및

상기 수집 대상 데이터를 분산 처리하는 상기 복수의 선택 데이터 노드를 포함하되,

상기 복수의 데이터 노드 각각으로서 동작하는 사용자 장치는 예측 데이터 통신량 및 예측 프로세서 사용량에 대한 정보를 고려하여 데이터 처리 가능 정보를 전송하고,

상기 예측 데이터 통신량은 상기 사용자 장치에서 구동되는 어플리케이션에서 사용할 것으로 예측되는 데이터의 크기이고,

상기 예측 프로세서 사용량은 상기 어플리케이션에서 사용할 것으로 예측되는 프로세서의 사용량이고,

상기 데이터 처리 가능 정보의 전송 주기는 상기 예측 데이터 통신량 및 상기 예측 프로세서 사용량의 변화를 기반으로 설정되는 것을 특징으로 하는 사용자 장치 기반의 빅데이터 분산 처리 시스템.

청구항 7

제6항에 있어서,

상기 복수의 데이터 노드는 상기 마스터 노드로 데이터 처리 가능 정보를 전송하도록 구현되고,

상기 데이터 처리 가능 정보는 상기 사용자 장치에 대한 사용자의 사용 특성에 대한 정보를 더 포함하는 것을 특징으로 하는 사용자 장치 기반의 빅데이터 분산 처리 시스템.

청구항 8

제7항에 있어서,

상기 데이터 처리 가능 정보는 상기 데이터 노드의 통신 상태, CPU(central processing unit) 사용률, 메모리 사용률에 대한 정보를 포함하는 것을 특징으로 하는 사용자 장치 기반의 빅데이터 분산 처리 시스템.

청구항 9

제8항에 있어서,

상기 선택 데이터 노드는 상기 마스터 노드로부터 수신한 푸시 알림 정보에 포함된 데이터 처리 지시를 기반으로 할당 데이터를 상기 수집 서버로부터 수신하여 처리하고 처리 결과를 서비스 데이터베이스로 전송하는 것을 특징으로 하는 사용자 장치 기반의 빅데이터 분산 처리 시스템.

청구항 10

제9항에 있어서,

상기 마스터 노드는 특정 시간까지 상기 할당 데이터가 처리되지 않은 미처리 데이터 노드를 파악하고, 상기 미처리 데이터 노드로 푸시 알림으로 재요청하거나 또는 상기 미처리 데이터 노드의 임무 취소 후 다른 데이터 노드에게 해당 임무를 지시하는 것을 특징으로 하는 사용자 장치 기반의 빅데이터 분산 처리 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 빅데이터 처리에 관한 것으로서, 보다 상세하게는 사용자 장치 기반의 빅데이터 분산 처리 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 2012년 미국의 리서치 전문업체인 가트너 (Gartner)는 빅데이터를 ‘데이터 양 (volume)이 많고, 데이터 형태가

다양(variety)하고, 데이터 속도(velocity)가 빠른 데이터' 로 정의하였다. 위키피디아(wikipedia)에서는 빅데이터를 기존 데이터베이스 관리 도구로서 데이터 수집, 저장, 관리, 분석 영역을 넘어선 대량의 정형 또는 비정형 데이터 집합' 으로 정의하였다.

[0003] 모바일 회사인 Tech Spartan에 의하면 2014년 기준 인터넷에서 1분 동안 생산되는 데이터량은 페이스북에서는 600,000건의 로그인이 되고, 트위터에서는 433,000건의 메시지가 트윗 (tweet)되고, 구글에서는 419만건의 검색이 이루어지고, 유튜브에서는 306시간 분량의 비디오가 업로드되고, 애플 아이튠즈에서는 50,200개 앱이 다운로드되고, 아마존에서는 80,000건 거래가 이루어지고, 이메일은 136,319,444개 전송된다고 하였다. 오늘날, 이처럼 대용량 데이터 시대를 맞이하여 IBM, HP, SAP, MS, EMC, 오라클 같은 대형 IT 벤더 회사들은 빅데이터 솔루션을 개발 중에 있다.

[0004] 빅데이터 시대가 도래하면서 새로운 처리/분석 패러다임이 요구되고 있고 미국중심으로 통계엔진인 R이 기업용 분석 플랫폼으로 확산되고 있다. R은 이미 구글, 페이스북, 야후, 아마존 등 닷컴 기업의 분석 플랫폼으로 사용중에 있다. 특히, 오라클, IBM, 테라데이터 등 빅데이터의 고성능 분석을 추구하는 IT업체들은 실시간 분석을 위해 메모리나 데이터베이스에서 직접 분석을 실시하는 인-메모리 분석(in-memory analytics) 혹은 인-데이터베이스 분석 (in-database analytics)을 할 수 있는 고성능 컴퓨팅 (high performance computing; HPC) 시스템에 R을 기본분석 플랫폼으로 채택하고 있다. R은 위와 같이 많은 활용에도 불구하고 확장성 (scalability)이 떨어지는 단점을 갖고 있다.

[0005] 따라서, R의 기본패키지로는 제한된 데이터 규모에서 만이 처리되고 작동된다.

[0006] R의 사용자들은 대용량 데이터 처리를 위해 여러 가지 방법을 제시하였는데 그 중에서 ff 패키지와 bigmemory 패키지 사용은 기존의 모든 데이터를 메모리에 로딩 후 처리하는 작업방식에서 벗어나 데이터 구조만을 메모리에 로딩하여 처리하는 방식으로 너무 느리다는 단점과 또한 물리적 메모리 확장의 한계로 인해 대용량 데이터 처리에 한계를 갖고 있다.

[0007] Hadoop은 대용량 데이터를 분산처리 할 수 있는 오픈 소스 플랫폼이다. R과 Hadoop의 통합환경으로는 Rhipe과 Rhadoop 등이 있다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 10-2013-0169007

발명의 내용

해결하려는 과제

[0009] 본 발명의 일 측면은 사용자 장치 기반의 빅데이터 분산 처리 방법을 제공한다.

[0010] 본 발명의 다른 측면은 사용자 장치 기반의 빅데이터 분산 처리 방법을 수행하는 장치를 제공한다.

과제의 해결 수단

[0011] 본 발명의 일 측면에 따른 사용자 장치 기반의 빅데이터 분산 처리 방법은, 수집 서버가 수집 대상 데이터를 분할 적재하는 단계, 상기 수집 서버는 마스터 노드에게 상기 수집 대상 데이터의 저장 상태를 전송하는 단계, 상기 마스터 노드가 복수의 데이터 노드의 상태 정보를 기반으로 상기 수집 대상 데이터의 처리를 상기 복수의 데이터 노드 중 복수의 선택 데이터 노드로 요청하는 단계와 상기 복수의 선택 데이터 노드가 상기 수집 대상 데이터를 분산 처리하는 단계를 포함할 수 있다.

[0012] 한편, 사용자 장치 기반의 빅데이터 분산 처리 방법은 상기 복수의 데이터 노드는 상기 마스터 노드로 데이터 처리 가능 정보를 전송하는 단계를 더 포함할 수 있다.

[0013] 또한, 상기 데이터 처리 가능 정보는 상기 데이터 노드의 통신 상태, CPU(central processing unit) 사용률, 메모리 사용률에 대한 정보를 포함할 수 있다.

[0014] 또한, 상기 선택 데이터 노드는 상기 마스터 노드로부터 수신한 푸시 알림 정보에 포함된 데이터 처리 지시를 기반으로 할당 데이터를 상기 수집 서버로부터 수신하여 처리하고 처리 결과를 서비스 데이터베이스로 전송할 수 있다.

[0015] 또한, 상기 마스터 노드는 특정 시간까지 상기 할당 데이터가 처리되지 않은 미처리 데이터 노드를 파악하고, 상기 미처리 데이터 노드로 푸시 알림으로 재요청하거나 또는 상기 미처리 데이터 노드의 임무 취소 후 다른 데이터 노드에게 해당 임무를 지시할 수 있다.

[0016] 본 발명의 또 다른 측면에 따른 사용자 장치 기반의 빅데이터 분산 처리 시스템은 수집 대상 데이터를 분할 적재하는 단계; 마스터 노드에게 상기 수집 대상 데이터의 저장 상태를 전송하도록 구현되는 수집 서버, 복수의 데이터 노드의 상태 정보를 기반으로 상기 수집 대상 데이터의 처리를 상기 복수의 데이터 노드 중 복수의 선택 데이터 노드로 요청하도록 구현되는 마스터 노드와 상기 수집 대상 데이터를 분산 처리하는 상기 복수의 선택 데이터 노드를 포함할 수 있다.

[0017] 한편, 상기 복수의 데이터 노드는 상기 마스터 노드로 데이터 처리 가능 정보를 전송하도록 구현될 수 있다.

[0018] 또한, 상기 데이터 처리 가능 정보는 상기 데이터 노드의 통신 상태, CPU(central processing unit) 사용률, 메모리 사용률에 대한 정보를 포함할 수 있다.

[0019] 또한, 상기 선택 데이터 노드는 상기 마스터 노드로부터 수신한 푸시 알림 정보에 포함된 데이터 처리 지시를 기반으로 할당 데이터를 상기 수집 서버로부터 수신하여 처리하고 처리 결과를 서비스 데이터베이스로 전송할 수 있다.

[0020] 또한, 상기 마스터 노드는 특정 시간까지 상기 할당 데이터가 처리되지 않은 미처리 데이터 노드를 파악하고, 상기 미처리 데이터 노드로 푸시 알림으로 재요청하거나 또는 상기 미처리 데이터 노드의 임무 취소 후 다른 데이터 노드에게 해당 임무를 지시할 수 있다.

발명의 효과

[0021] 본 발명의 실시예에 따른 사용자 장치 기반의 빅데이터 분산 처리 방법 및 장치에 따르면, 이미 대중화되고 PC(personal computer) 급 컴퓨팅 능력을 갖춰 나가고 있는 개인용 모바일 스마트 디바이스(스마트폰 등)의 컴퓨팅 자원(CPU(central processing unit), 메모리(mempry) 등)을 공유(sharing) 함으로써 수십 또는 수백대 분량의 빅데이터 처리를 위한 데이터 노드(서버)도입을 대체 할 수 있다. 따라서, 엔터프라이즈 측면에서는 대규모 비용 절감이 가능해질 수 있고, 빅데이터 처리 시스템 도입 진입 장벽이 낮추어질 수 있다.

[0022] 또한 모바일 자원을 기업에게 제공하는 개인의 측면에서는 대부분의 시간을 유휴 상태로 대기 중인 개인의 스마트 디바이스의 자원을 제공하고 그 대가로 기업으로부터 다양한 콘텐츠 서비스를 무상으로 제공받거나, 디바이스 제공 수수료와 같은 금전적 수익을 제공받을 수도 있다. 즉, 대용량 데이터 분산 처리 분야에 있어, 개인과 기업 모두 쌍방향 이익을 도모할 수 있는 또 한가지의 새로운 IT 관련 솔루션이 될 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 실시예에 따른 사용자 장치를 기반으로 한 빅데이터 분산 처리 시스템을 나타낸 개념도이다.

도 2는 본 발명의 실시예에 따른 사용자 장치를 기반으로 한 빅데이터 분산 처리 방법을 나타낸 순서도이다.

도 3은 본 발명의 실시예에 따른 빅데이터 분산 처리 시나리오를 나타낸 개념도이다.

도 4는 본 발명의 실시예에 따른 데이터 분산 처리 방법을 나타낸 개념도이다.

발명을 실시하기 위한 구체적인 내용

[0024] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예와 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든

범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조 부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0025] 이하, 도면들을 참조하여 본 발명의 바람직한 실시예들을 보다 상세하게 설명하기로 한다.

[0027] 외국 IT(information technology) 회사가 개발하여 현재 기업 및 국가 기관에서 주로 사용 중인 HADOOP 빅데이터 분산 저장 및 처리 시스템 또는 이와 유사한 현대의 빅데이터 처리를 위한 솔루션들은 분산 데이터 처리를 위해 많은 노드(서버 컴퓨터)를 해당 기업에서 직접 구매하거나 임대하여 사용하는 방식을 사용한다.

[0028] 이러한 방식은 빅데이터 처리를 위한 시스템의 초기 구축 및 유지 보수를 위해 막대한 자금이 투자될 뿐만 아니라, 처리할 데이터가 없어 유휴 상태일 때도 전원 공급 및 장비 공간에 대한 임대료 등이 고정비로 발생하는 등 해당 시스템을 도입하기 위한 다양한 진입 장벽이 존재한다.

[0029] 본 발명의 실시예에 따른 사용자 장치 기반의 빅데이터 분산 처리 방법에서는 모바일 스마트 기기를 활용한 빅데이터 처리 방법이 제안된다. 이미 대중화되고 PC(personal computer) 급 컴퓨팅 능력을 갖춰 나가고 있는 개인용 모바일 스마트 디바이스(스마트폰 등)의 컴퓨팅 자원(CPU(central processing unit), 메모리(mempry) 등)을 공유(sharing) 함으로써 수십 또는 수백대 분량의 빅데이터 처리를 위한 데이터 노드(서버)도입을 대체할 수 있다. 따라서, 엔터프라이즈 측면에서는 대규모 비용 절감이 가능해질 수 있고, 빅데이터 처리 시스템 도입 진입 장벽이 낮추어질 수 있다.

[0030] 또한 모바일 자원을 기업에게 제공하는 개인의 측면에서는 대부분의 시간을 유휴 상태로 대기 중인 개인의 스마트 디바이스의 자원을 제공하고 그 대가로 기업으로부터 다양한 콘텐츠 서비스를 무상으로 제공받거나, 디바이스 제공 수수료와 같은 금전적 수익을 제공받을 수도 있다.

[0031] 본 발명의 실시예에 따른 사용자 장치 기반의 빅데이터 분산 처리 방법이 사용됨으로써 대용량 데이터 분산 처리 분야에 있어, 개인과 기업 모두 쌍방향 이익을 도모할 수 있는 또 한가지의 새로운 IT 관련 솔루션이 되는 것을 그 목표로 한다.

[0033] 도 1은 본 발명의 실시예에 따른 사용자 장치를 기반으로 한 빅데이터 분산 처리 시스템을 나타낸 개념도이다.

[0034] 도 1에서는 사용자 장치를 기반으로 빅데이터를 분산 처리하기 위한 빅데이터 분산 처리 시스템이 개시된다.

[0035] 빅데이터 분산 처리 시스템은 데이터 노드(100), 마스터 노드(120), 서비스 데이터베이스(160), 수집 서버를 포함할 수 있다.

[0036] 데이터 노드(100)는 사용자 장치일 수 있다. 사용자 장치는 사용자들이 가지고 다니는 스마트폰과 같은 장치일 수 있다. 데이터 노드(100)는 빅데이터 분산 처리에 동의한 사용자 장치로서 수집 서버(120), 마스터 노드(140)와 통신 가능한 스마트 어플리케이션을 설치하고, 데이터 노드(100)로서 등록된 장치일 수 있다.

[0037] 데이터 노드(100)는 주기적으로 마스터 노드(140)와 통신하면서 기기 상태를 전송할 수 있고, 마스터 노드(140)는 주기적으로 데이터 노드(100)의 상태 확인 및 관련 정보의 갱신을 수행할 수 있다.

[0038] 마스터 노드(140)는 수집 서버(120), 데이터 노드(100)를 관리하기 위한 노드일 수 있다. 또한, 마스터 노드(140)는 분석 처리 알고리즘을 관리하고, 분석 처리 스케줄을 관리할 수 있다.

[0039] 마스터 노드(140)는 등록된 데이터 노드(100) 중 자원 사용률이 낮은 노드에게 빅데이터 처리에 필요한 수집 서버 정보, 데이터셋 정보, 처리 알고리즘을 푸시 알림을 전송하여 데이터 처리를 지시할 수 있다. 데이터 노드(100)는 마스터 노드(140)로부터 수신한 푸시 알림 정보에 포함된 데이터 처리 지시를 바탕으로 자신에게 할당된 소량의 데이터를 수집 서버(120)로부터 수신하여 처리하고 결과를 서비스 데이터베이스(160)에 저장할 수 있다.

[0040] 수집 서버(120)는 내부 대용량 데이터, 외부 수집 대상 데이터를 수집할 수 있다. 수집 서버(120)는 복수개일 수 있다. 내부 대용량 데이터는 다양한 통신 프로토콜(예를 들어, FTP(file transfer protocol)/SFTP(secure FTP), NFS)를 기반으로 수집될 수 있다. 예를 들어, 내부 대용량 데이터는 각종 지적도 데이터, 크롤링된 비정형 데이터, 통신 정보, 카드 사용 정보, 사진 이미지 데이터, 대용량 빅데이터 등을 포함할 수 있다. 외부 수집 대상 데이터는 통신 프로토콜(예를 들어, FTP/SFTP, SMTP, HTTP/HTTPS, OFF LINE 수집)을 기반으로 수집될 수 있다. 예를 들어, 외부 수집 대상 데이터는 실시간 운행 기록 정보, IoT(internet of things) 센서 이벤트 정보, 기타 대용량 데이터 등을 포함할 수 있다.

[0041] 수집 서버(120)는 다양한 통신 프로토콜(FTP/HTTP/NFS 등)을 통해 수집 서버(120)로 데이터를 분할 적재 후 마스터 노드(140)에게 저장 상태를 전송할 수 있다.

[0042] 서비스 데이터베이스(160)는 마스터 노드(140)로부터 데이터 노드(100)의 데이터 처리 결과를 수신할 수 있고, 수신한 데이터 처리 결과는 통계 분석 자료, 시각화 자료, 의사 결정 자료로서 사용될 수 있다.

[0044] 도 2는 본 발명의 실시예에 따른 사용자 장치를 기반으로 한 빅데이터 분산 처리 방법을 나타낸 순서도이다.

[0045] 도 2에서는 데이터 노드로 설정된 사용자의 사용자 장치를 기반으로 빅데이터를 분산 처리하기 위한 방법이 개시된다.

[0046] 도 2를 참조하면, 사용자 장치는 빅데이터의 분산 처리를 위한 분산 처리 어플리케이션을 설치할 수 있다.

[0047] 사용자 장치는 분산 처리 어플리케이션을 설치하고 빅데이터의 분산 처리에 동의하고, 수집 서버, 마스터 노드와 통신할 수 있다(단계 S200).

[0048] 사용자 장치는 분산 처리 어플리케이션을 탑재하고 마스터 노드에 데이터 노드로서 등록될 수 있다.

[0049] 수집 대상 데이터가 수집 서버로 분할 적재되고, 수집 서버는 마스터 노드에게 저장 상태를 전송한다(단계 S210).

[0050] 내부/외부 수집 대상 데이터는 FTP/HTTP/NFS 등 다양한 통신 프로토콜을 기반으로 수집 서버로 데이터를 분할 적재 후 마스터 노드에게 저장 상태 전송할 수 있다(단계 S220).

[0051] 예를 들어, 내부 대용량 데이터는 각종 지적도 데이터, 크롤링된 비정형 데이터, 통신 정보, 카드 사용 정보, 사진 이미지 데이터, 대용량 빅데이터 등을 포함할 수 있다. 외부 수집 대상 데이터는 통신 프로토콜(예를 들어, FTP/SFTP, SMTP, HTTP/HTTPS, OFF LINE 수집)을 기반으로 수집될 수 있다. 예를 들어, 외부 수집 대상 데이터는 실시간 운행 기록 정보, IoT(internet of things) 센서 이벤트 정보, 기타 대용량 데이터 등을 포함할 수 있다.

[0052] 데이터 노드는 주기적으로 마스터 노드와 통신하며 기기 상태를 전송하고, 마스터 노드는 주기적으로 데이터 노드의 상태를 확인하고 관련 정보를 갱신할 수 있다(단계 S230).

[0053] 마스터 노드는 등록된 데이터 노드 중 자원 사용률이 낮은 데이터 노드에게 빅데이터 처리에 필요한 수집 서버 정보, 데이터셋 정보, 처리 알고리즘을 푸시 알림을 전송하여 데이터 처리를 지시할 수 있다(단계 S240).

[0054] 데이터 노드는 마스터 노드로부터 수신한 푸시 알림 정보에 포함된 데이터 처리 지시를 바탕으로 자신에게 할당된 소량의 데이터를 수집 서버로부터 수신하여 처리하고 결과를 서비스 데이터베이스에 저장할 수 있다(단계 S250).

[0055] 즉, 사용자 장치 기반의 빅데이터 분산 처리 방법은, 수집 서버가 수집 대상 데이터를 분할 적재하는 단계, 수집 서버는 마스터 노드에게 수집 대상 데이터의 저장 상태를 전송하는 단계, 마스터 노드가 복수의 데이터 노드의 상태 정보를 기반으로 수집 대상 데이터의 처리를 복수의 데이터 노드 중 복수의 선택 데이터 노드로 요청하는 단계, 복수의 선택 데이터 노드가 수집 대상 데이터를 분산 처리하는 단계를 포함할 수 있다.

[0056] 선택 데이터 노드는 마스터 노드로부터 수신한 푸시 알림 정보에 포함된 데이터 처리 지시를 기반으로 할당 데이터를 수집 서버로부터 수신하여 처리하고 처리 결과를 서비스 데이터베이스로 전송할 수 있다.

[0057] 마스터 노드는 특정 시간까지 할당 데이터가 처리되지 않은 미처리 데이터 노드를 파악하고, 미처리 데이터 노드로 푸시 알림으로 재요청하거나 또는 미처리 데이터 노드의 임무 취소 후 다른 데이터 노드에게 해당 임무를 지시할 수 있다.

[0059] 도 3은 본 발명의 실시예에 따른 빅데이터 분산 처리 시나리오를 나타낸 개념도이다.

[0060] 도 3에서는 빅데이터 분산 처리 시나리오가 개시된다.

[0061] 도 3을 참조하면, 데이터 노드(사용자 장치)의 준비 단계가 수행될 수 있다(단계 S300).

[0062] 예를 들어, 복수개의 스마트폰(예를 들어, 100대)에 데이터 자원을 관리하거나 계산하기 위한 분산 처리 어플리케이션이 설치될 수 있다.

[0063] 데이터 노드(스마트폰)는 앱 내에 포함된 각종 설정 정보를 기반으로 마스터 노드에게 데이터 처리 가능 정보를

주기적으로 전송할 수 있다. 데이터 처리 가능 정보는 단말기 통신 상태(와이파이, 무선 통신), CPU, 메모리 사용률 등 데이터 처리 가능 여부 상태 정보일 수 있다.

[0064] 마스터 노드는 데이터 처리 가능 정보에 포함된 데이터 노드의 상태를 분석하여 가용 데이터 노드 정보, 개수, 용량 등에 대한 정보를 관리할 수 있다.

[0065] 데이터 수집 단계가 수행될 수 있다(단계 S310).

[0066] 예를 들어, 외부에서 웹 크롤링을 통해 수집 서버로 지난 10년간 신문 기사 텍스트 데이터 수집하고 데이터명을 신문 데이터셋(예를 들어, 용량 10GB)으로 가정할 수 있다.

[0067] 수집 서버는 현재 수집된 신문 데이터 셋을 정상적으로 수신 및 저장했다고 마스터 노드에게 수집 관련 정보 전송할 수 있다. 수집 관련 정보는 수집 서버명, 파일명, 경로, 용량, 수집 경로 등에 대한 정보를 포함할 수 있다.

[0068] 마스터 노드는 메모리에 상주된 데이터 관리 프레임에 해당 내용을 저장하거나 갱신하고 수집 노드 관리 데이터베이스에도 물리적으로 저장 및 갱신을 수행할 수 있다.

[0069] 분산 처리를 위한 임무 할당 단계가 수행될 수 있다(단계 S320).

[0070] 예를 들어, 사용자(프로그래머)는 특정 목적에 의한 신속한 데이터 분산 처리를 위해 마스터 노드에게 신문 데이터 셋에서 '대한민국' 단어의 개수를 카운트하라고 지시할 수 있다. 마스터 노드는 현재 관리 중인 데이터 노드(스마트폰)에서 자원을 가용할 수 있다고 판단되는 데이터 노드를 계산할 수 있다.

[0071] 신문 데이터 셋이 저장된 수집 서버 정보와 신문 데이터셋(10G)에서 데이터 노드별로 얼마의 데이터를 할당할지에 대한 데이터 조각 정보, 각 데이터 노드에서 수행할 작업 내용(대한민국 단어 카운트 지시)등을 배열에 포함하여 푸시 알림 콘텐츠에 포함하여 데이터 노드로 전송할 수 있다.

[0072] 빅데이터 분산 처리 단계가 수행될 수 있다(단계 S330).

[0073] 작업 내용을 푸시 알림으로 수신한 데이터 노드는 푸시 알림에 포함된 데이터 노드의 데이터 처리 할당 정보를 기반으로 수집 서버와 통신하여 자신에게 할당된 신문 데이터 셋의 특정 구간 데이터를 요청할 수 있다.

[0074] 수집 서버는 데이터 노드로부터 요청 받은 정보를 바탕으로 신문 데이터 셋의 정보를 분할하여 각 데이터 노드들에게 전송할 수 있다.

[0075] 데이터 노드 100대는 수집 서버로부터 자신의 할당량 100 메가 바이트씩을 분할하여 수신하고 해당 데이터 조각에서 대한민국 단어 카운트 수행할 수 있다.

[0076] 각 데이터 노드는 결과값(예를 들어, 대한민국 : 100 count)을 수행 결과 정보에 저장하고 마스터 노드에게 성공 여부 및 결과 값을 전송할 수 있다.

[0077] 마스터 노드로부터 결과값을 정상적으로 수신했다는 정보를 수신한 데이터 노드는 자신이 보유한 데이터 조각을 삭제하여 용량을 확보할 수 있다.

[0078] 마스터 노드는 특정 시간까지 성공 여부 결과가 도착하지 않은 데이터 노드를 파악하고 관리할 수 있다. 예를 들어, 마스터 노드는 푸시 알림으로 재요청하거나 또는 임무 취소 후 다른 노드에게 해당 임무를 지시할 수 있다.

[0079] 모든 데이터 노드로부터 결과값을 수신했을 경우, 전체 데이터 노드에서 도착한 결과 값(예를 들어, 대한민국 : 10000 count)을 취합할 수 있다. 취합된 결과값을 서비스 데이터베이스 해당 임무 결과에 저장하고 사용자(프로그래머)에게 결과값을 반환할 수 있다.

[0081] 도 4는 본 발명의 실시예에 따른 데이터 분산 처리 방법을 나타낸 개념도이다.

[0082] 도 4에서는 사용자 장치의 빅데이터 분산 처리 여부를 결정하기 위한 방법이 개시된다.

[0083] 도 4를 참조하면, 사용자 장치는 빅데이터 분산 처리를 위해 자신의 컴퓨팅 자원을 소모해야 한다. 따라서, 사용자 장치는 현재 사용자 장치의 상태가 어떠한지 여부를 판단하여 빅데이터 분산 처리를 위한 컴퓨팅 자원의 대여 여부를 결정할 수 있다.

[0084] 사용자가 현재 사용자 장치를 통해 임계치 이상의 컴퓨팅 자원을 사용하고 있는 경우가 가정될 수 있다. 예를

들어, 사용자가 일정 주기 내에 임계값 이상의 데이터 통신을 수행하고, 임계치 이상의 사용자 장치의 프로세서를 사용하는 경우가 가정될 수 있다.

[0085] 이러한 경우, 사용자 장치는 데이터 처리 가능 정보를 전송시 일정 주기 내의 데이터 통신량과 프로세서의 사용량을 고려하여 데이터 처리량을 결정하여 전송할 수 있다.

[0086] 사용자 장치는 일정 주기 내에 수행될 사용자 장치의 데이터 사용량/프로세서의 사용량을 예측할 수 있다. 제1 어플리케이션이 사용자 장치에서 실행 중인 경우, 사용자 장치 상에서 제1 어플리케이션이 앞으로 사용할 예측 데이터량, 예측 프로세싱량을 결정할 수 있다. 제1 어플리케이션이 계속적으로 임계 범위 내의 데이터량/일정한 프로세싱량을 사용할 경우, 예측 데이터량(400)/예측 프로세싱량(420)을 기존의 데이터량, 프로세싱량의 평균을 기반으로 일정한 양으로 예측하여 결정할 수 있다. 반대로, 제1 어플리케이션이 변화량이 큰 데이터량/ 프로세싱량을 사용할 경우, 예측 값을 기존의 데이터양, 프로세싱 양의 최대값을 기반으로 예측하여 결정할 수 있다.

[0087] 사용자 장치는 사용자 장치에서 사용되는 데이터량, 프로세싱량의 변화를 기반으로 데이터 처리 가능 정보의 전송 주기를 결정할 수도 있다. 제1 어플리케이션이 계속적으로 일정한 데이터량/일정한 프로세싱량을 사용할 경우, 데이터 처리 가능 정보의 전송 주기를 상대적으로 길게 설정할 수 있다. 반대로, 제1 어플리케이션이 변화량이 큰 데이터량/프로세싱량을 사용할 경우, 데이터 처리 가능 정보의 전송 주기를 상대적으로 짧게 설정하여 제1 어플리케이션에서 사용되는 데이터량/프로세싱량의 변화를 보다 자주 전송할 수 있다.

[0088] 또한, 본 발명의 실시예에 따르면, 사용자 별로 데이터/프로세서 사용 특성을 추가적으로 고려하여 데이터 처리 가능 정보가 생성될 수도 있다. 분산 처리 어플리케이션은 사용자의 사용자 장치 사용 특성을 결정하고, 데이터 처리 가능 정보를 생성할 수도 있다. 예를 들어, 특정 사용자의 경우, 사용자 장치 상에서 동시에 다양한 어플리케이션(또는 프로세싱 성능을 많이 사용하는 어플리케이션)을 사용하는 사용 패턴을 가지고 다른 사용자의 경우, 어플리케이션을 거의 사용하지 않고, 사용한다고 해도 일정량 이하의 프로세싱 성능을 사용하는 어플리케이션만 사용할 수 있다. 이러한 사용자의 사용자 장치 사용 특성 정보는 분산 처리 어플리케이션에 기록되어 있을 수 있고, 분산 처리 어플리케이션은 이러한 사용자의 사용 특성을 기반으로 데이터 처리 가능 정보를 생성할 수도 있다.

[0090] 전술한 방법은 애플리케이션으로 구현되거나 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다.

[0091] 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들일 수 있고, 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다.

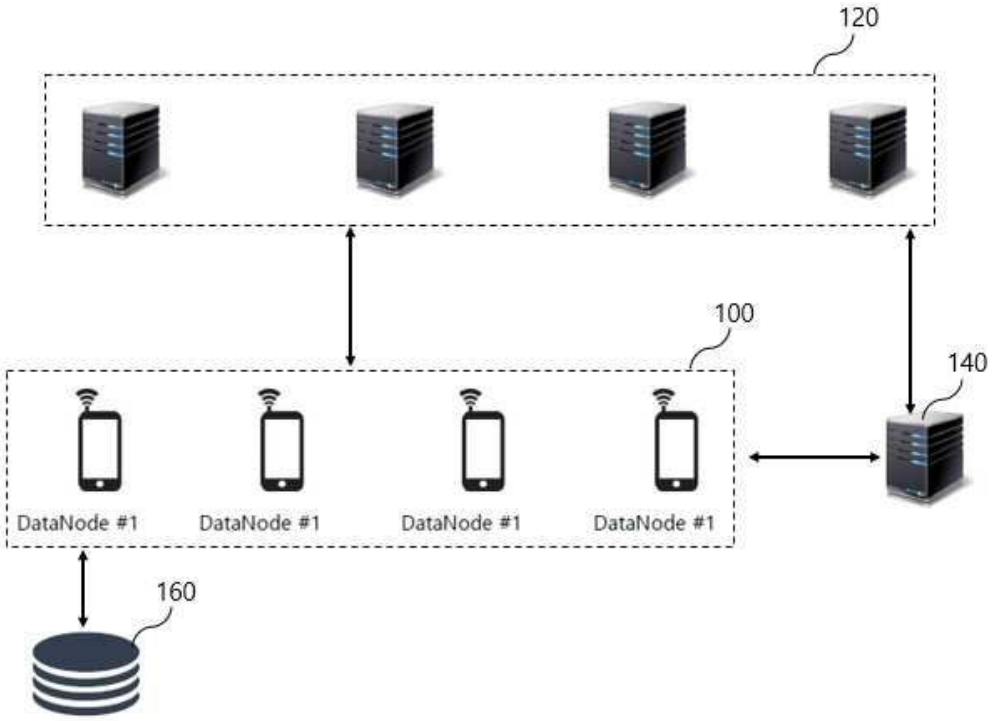
[0092] 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD 와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0093] 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

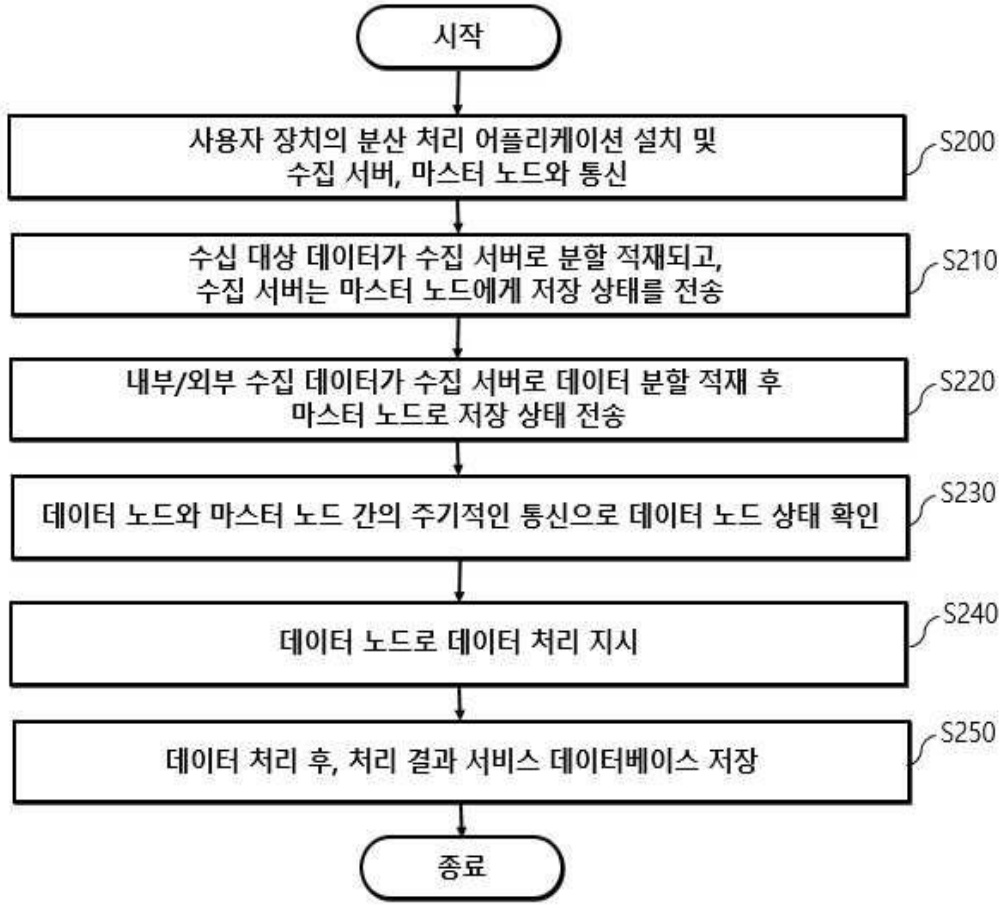
[0094] 이상에서는 실시예들을 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

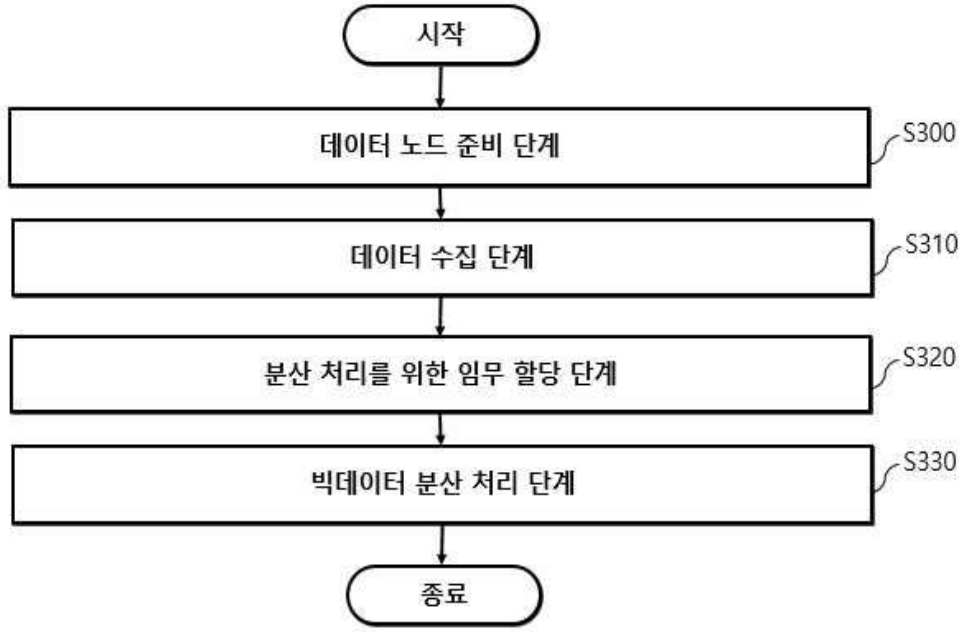
도면1



도면2


시작
사용자 장치의 분산 처리 어플리케이션 설치 및
수집 서버, 마스터 노드와 통신
S200
수집 대상 데이터가 수집 서버로 분할 적재되고,
수집 서버는 마스터 노드에게 저장 상태를 전송
S210
내부/외부 수집 데이터가 수집 서버로 데이터 분할 적재 후
마스터 노드로 저장 상태 전송
S220
데이터 노드와 마스터 노드 간의 주기적인 통신으로 데이터 노드 상태 확인
S230
데이터 노드로 데이터 처리 지시
S240
데이터 처리 후, 처리 결과 서비스 데이터베이스 저장
S250
종료

도면3


시작
데이터 노드 준비 단계
S300
데이터 수집 단계
S310
분산 처리를 위한 임무 할당 단계
S320
빅데이터 분산 처리 단계
S330
종료

도면4

