(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0118357
(43) 공개일자 2010년11월05일

(51) Int. Cl.
G06F 9/50 (2006.01) *G06F 9/46* (2006.01)
G06F 9/00 (2006.01)
(21) 출원번호 10-2009-0037151
(22) 출원일자 2009년04월28일
심사청구일자 2009년04월28일

(71) 출원인
인하대학교 산학협력단
인천 남구 용현동 253 인하대학교
(72) 발명자
이종식
인천시 남구 용현3동 인하대학교 하이테크관 1409호
노창현
인천시 남구 용현3동 인하대학교 하이테크관 1407호
김태영
인천시 남구 용현3동 인하대학교 하이테크센터 1407호
(74) 대리인
특허법인무한

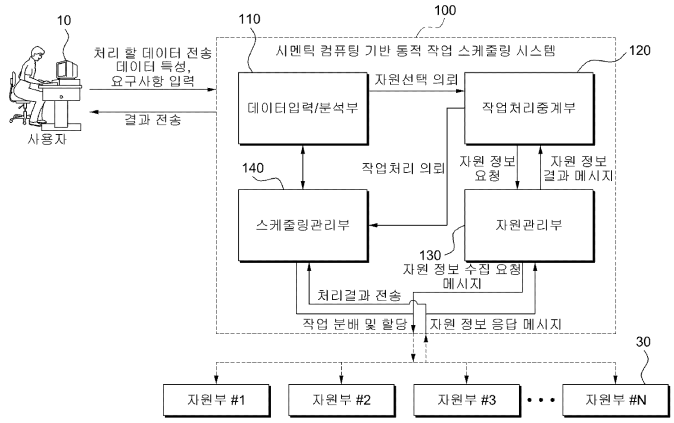
전체 청구항 수 : 총 5 항

(54) 분산 처리를 위한 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템

(57) 요 약

분산 처리를 위한 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템을 제공한다. 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 작업 대상인 데이터 및, 상기 데이터에 해당하는 데이터 특성과 상기 데이터를 처리하기 위한 요구사항을 입력 받는 데이터 입력/분석부; 상기 데이터가 처리되도록 전반의 동작을 제어하는 작업처리 중계부; 적어도 하나의 자원과 통신을 통하여 상기 자원의 속성 정보를 수집 및 유지하고 상기 데이터 특성과 요구사항에 대응되는 가용 자원을 선별하는 자원 관리부; 및, 상기 자원과의 통신을 통하여 상기 선별된 가용 자원 중 상기 데이터를 처리할 자원을 선택하여 상기 선택된 자원에 상기 데이터를 분배하고 상기 분배된 데이터에 대한 처리 결과를 제공하는 스케줄링 관리부를 포함한다.

대 표 도



특허청구의 범위

청구항 1

작업 대상인 데이터 및, 상기 데이터에 해당하는 데이터 특성과 상기 데이터를 처리하기 위한 요구사항을 입력받는 데이터 입력/분석부;

상기 데이터가 처리되도록 전반의 동작을 제어하는 작업처리 중계부;

적어도 하나의 자원과 통신을 통하여 상기 자원의 속성 정보를 수집 및 유지하고 상기 데이터 특성과 요구사항에 대응되는 가용 자원을 선별하는 자원 관리부; 및,

상기 자원과의 통신을 통하여 상기 선별된 가용 자원 중 상기 데이터를 처리할 자원을 선택하여 상기 선택된 자원에 상기 데이터를 분배하고 상기 분배된 데이터에 대한 처리 결과를 제공하는 스케줄링 관리부

를 포함하는 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템.

청구항 2

제1항에 있어서,

상기 데이터 특성은 데이터의 종류, 데이터의 크기 중 적어도 하나를 포함하고,

상기 요구 사항은 상기 데이터를 처리하는데 사용할 예산 금액, 상기 데이터를 처리해야 하는 마감 기한 중 적어도 하나를 포함하고,

상기 데이터 입력/분석부는,

상기 입력된 데이터 특성을 정수 형태로 변환한 후 상기 작업처리 중계부로 전달하는 데이터 분석기와,

상기 입력된 요구 사항을 정수 형태로 변환한 후 상기 작업처리 중계부로 전달하는 요구사항 분석기를 포함하는, 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템.

청구항 3

제1항에 있어서,

상기 작업처리 중계부는,

상기 자원 관리부에 상기 가용 자원에 대한 자원 목록을 요청하는 가용자원 요청기와,

상기 자원 목록의 요청에 따라 상기 자원 관리부에서 제공하는 상기 가용 자원의 자원 목록을 수신하는 가용자원 수신기와,

상기 데이터 특성과 요구사항에 대하여 정의된 추론 규칙을 이용하여 상기 수신된 자원 목록의 가용 자원 중 상기 데이터 특성과 요구사항에 대응되는 자원을 선택하는 자원 선택기를 포함하는, 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템.

청구항 4

제1항에 있어서,

상기 자원 관리부는,

상기 자원으로 속성 정보 요청 메시지를 전송하는 자원정보 요청기와,

상기 요청 메시지에 대하여 상기 자원에서 제공하는 상기 속성 정보를 수신하는 자원정보 수신기와,

상기 자원을 관리하기 위해 상기 수신된 속성 정보를 온톨로지 표준 언어를 이용하여 표현하는 자원 온톨로지 관리기와,

상기 온톨로지 표준 언어로 표현된 속성 정보를 상기 자원과 연관하여 저장하는 자원정보 저장소와,

상기 작업처리 중계부의 제어 하에 상기 데이터 특성과 요구사항에 대응되는 가용 자원을 선별하여 상기 가용

자원으로 구성되는 자원 목록을 상기 작업처리 중계부로 제공하는 자원정보 송신기

를 포함하는, 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템.

청구항 5

제1항에 있어서,

상기 스케줄링 관리부는,

상기 자원의 동적인 속성 정보를 수신하는 자원상태정보 수신기와,

상기 자원의 속성 정보에 대하여 정의된 추론 규칙을 이용하여 상기 수신된 동적인 속성 정보에 따라 상기 가용 자원의 순위를 결정하는 추론 엔진부와,

상기 가용 자원 중 상기 등급이 가장 높은 자원에 상기 데이터를 분배 및 할당하는 작업 분배기와,

상기 데이터를 할당받은 자원으로부터 상기 데이터에 대한 처리 결과를 수신하여 수신된 처리 결과를 취합하는 처리완료 데이터 취합기를 포함하고,

상기 동적인 속성 정보는,

상기 자원의 CPU 사용 여부, CPU 사용률, 주기억 장치 사용률, 보조 기억 장치 사용률, 네트워크 상태 중 적어도 하나를 포함하는, 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템.

명 세 서

발명의 상세한 설명

기 술 분 야

[0001] 본 발명의 실시예들은 지리적으로 분산되어 있는 이기종의 자원에 작업을 분배하여 분산 처리하기 위한 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템에 관한 것이다.

배 경 기 술

[0002] 그리드 컴퓨팅(Grid computing)이나 분산 컴퓨팅(Distributed Computing) 환경에서 지리적으로 분산되어 있는 자원에 작업을 분배하는 방법으로는 "그리드에서 사용자의 QoS와 응용특성에 근거한 동적 자원 할당방법(국내 특허 등록 10-0621787)"이 있다. 상기한 종래의 기술은 그리드 컴퓨팅 환경에서 자원의 성능을 고려하여 단시간에 작업을 처리하기 위해 비선형모델을 선형 모델로 변환하여 목적함수를 도출한다. 그러나, 목적함수에 포함 된 인자로 자원의 계산 능력을 나타내는 CPU의 성능과 작업 부하량, 네트워크의 상태만을 표시하여 자원의 다양한 동적인 상태를 반영하여 효율적으로 스케줄링하기 어렵다. 또한, 새로운 종류의 자원을 정의하여 추가 및 갱신하거나 사용자의 요구사항과 처리 할 데이터의 특성에 따른 최적의 작업 스케줄링 기법을 적용하기에는 호환성과 확장성에 문제가 있다.

[0003] 기존의 지리적으로 분산되어 있는 자원에 작업을 분배하는 또 다른 방법으로는 "분산 처리 시스템의 작업 스케줄링 장치 및 방법(국내 공개 특허 특1998-079594)"이 있다. 상기한 종래의 기술은 임의의 주기에 따라 작업을 스케줄링하는 능력, 작업을 건너뛰는 능력, 원격으로 네트워크 내의 자원 상에서 발생하는 작업들을 스케줄링하는 방법을 제공한다. 그러나, 지리적으로 분산 된 컴퓨팅 환경에서 상기한 종래의 기술은 자원의 성능과 동적인 변화, 네트워크의 혼잡 등을 고려하지 않음으로써 자원 이용의 불균형을 초래하거나 네트워크의 혼잡을 초래할 수 있다.

발명의 내용

해결 하고자하는 과제

[0004] 본 발명은 이기종의 자원으로 이루어진 컴퓨팅 환경에서 보다 용이한 자원관리가 가능하고 시스템 확장이 용이한 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템을 제공한다.

[0005] 본 발명은 사용자의 예산과 데이터 특성에 따라 가장 적합한 자원을 이용하여 작업을 처리할 수 있는 시멘틱 컴

퓨팅 기반의 동적 작업 스케줄링 시스템을 제공한다.

[0006] 본 발명은 분산되어 있는 컴퓨팅 자원에 작업을 분배 및 할당하기 위한 효율적인 스케줄링 기법을 제공하는 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템을 제공한다.

과제 해결수단

[0007] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 작업 대상인 데이터 및, 상기 데이터에 해당하는 데이터 특성과 상기 데이터를 처리하기 위한 요구사항을 입력 받는 데이터 입력/분석부; 상기 데이터가 처리되도록 전반의 동작을 제어하는 작업처리 중계부; 적어도 하나의 자원과 통신을 통하여 상기 자원의 속성 정보를 수집 및 유지하고 상기 데이터 특성과 요구사항에 대응되는 가용 자원을 선별하는 자원 관리부; 및, 상기 자원과의 통신을 통하여 상기 선별된 가용 자원 중 상기 데이터를 처리할 자원을 선택하여 상기 선택된 자원에 상기 데이터를 분배하고 상기 분배된 데이터에 대한 처리 결과를 제공하는 스케줄링 관리부를 포함한다.

[0008] 본 발명의 일실시예에서 상기 데이터 입력/분석부는 상기 입력된 데이터 특성을 정수 형태로 변환한 후 상기 작업처리 중계부로 전달하는 데이터 분석기와, 상기 입력된 요구 사항을 정수 형태로 변환한 후 상기 작업처리 중계부로 전달하는 요구사항 분석기를 포함한다.

[0009] 본 발명의 일실시예에서 상기 작업처리 중계부는 상기 자원 관리부에 상기 가용 자원에 대한 자원 목록을 요청하는 가용자원 요청기와, 상기 자원 목록의 요청에 따라 상기 자원 관리부에서 제공하는 상기 가용 자원의 자원 목록을 수신하는 가용자원 수신기와, 상기 데이터 특성과 요구사항에 대하여 정의된 추론 규칙을 이용하여 상기 수신된 자원 목록의 가용 자원 중 상기 데이터 특성과 요구사항에 대응되는 자원을 선택하는 자원 선택기를 포함한다.

[0010] 본 발명의 일실시예에서 상기 자원 관리부는 상기 자원으로 속성 정보 요청 메시지를 전송하는 자원정보 요청기와, 상기 요청 메시지에 대하여 상기 자원에서 제공하는 상기 속성 정보를 수신하는 자원정보 수신기와, 상기 자원을 관리하기 위해 상기 수신된 속성 정보를 온톨로지 표준 언어를 이용하여 표현하는 자원 온톨로지 관리기와, 상기 온톨로지 표준 언어로 표현된 속성 정보를 상기 자원과 연관하여 저장하는 자원정보 저장소와, 상기 작업처리 중계부의 제어 하에 상기 데이터 특성과 요구사항에 대응되는 가용 자원을 선별하여 상기 가용 자원으로 구성되는 자원 목록을 상기 작업처리 중계부로 제공하는 자원정보 송신기를 포함한다.

[0011] 본 발명의 일실시예에서 상기 스케줄링 관리부는 상기 자원의 동적인 속성 정보를 수신하는 자원상태정보 수신기와, 상기 자원의 속성 정보에 대하여 정의된 추론 규칙을 이용하여 상기 수신된 동적인 속성 정보에 따라 상기 가용 자원의 순위를 결정하는 추론 엔진부와, 상기 가용 자원 중 상기 등급이 가장 높은 자원에 상기 데이터를 분배 및 할당하는 작업 분배기와, 상기 데이터를 할당받은 자원으로부터 상기 데이터에 대한 처리 결과를 수신하여 수신된 처리 결과를 취합하는 처리완료 데이터 취합기를 포함한다.

효 과

[0012] 본 발명의 일실시예에 따르면, 사용자의 요구 사항과 데이터의 특성에 따라 선택하는 자원들을 지식 기반의 규칙으로 정의하게 되는데, 여기서 정의된 규칙들은 사용자나 시스템 전문가에 의해 새롭게 추가, 수정, 삭제가 용이한 장점이 있으며 사용자의 예산과 데이터 특성에 따라 가장 적합한 자원들을 선택함으로써 마감기한 내에 처리하여 사용자의 만족도(QoS)를 높일 수 있다.

[0013] 본 발명의 일실시예에 따르면, 사용자가 처리하려는 데이터를 자원의 정적인 속성으로 평가하고 다양한 동적인 변화에 따른 여러 가지 스케줄링 기법을 지식 기반의 규칙으로 정의하여 처리함으로써 작업 처리 도중 발생할 수 있는 오류나 작업 손실에 대한 신속한 대처가 가능하여 효율적인 작업 처리를 제공할 수 있다. 또한, 낮은 작업 손실과 높은 작업 처리율 및 짧은 응답 시간을 제공함으로써 사용자 작업을 처리함에 있어서 높은 효율성 및 안정성을 제공 할 수 있다.

[0014] 본 발명의 일실시예에 따르면, 이기종의 자원으로 이루어진 컴퓨팅 환경에서 W3C(World Wide Web Consortium)에서 제정한 온톨로지 표준 언어인 OWL(Web Ontology Language)을 이용하여 자원들을 기술 및 관리함으로써 새로운 자원의 추가 및 삭제, 갱신 등 자원관리가 더욱 용이하고 시스템의 확장성과 호환성을 높일 수 있다.

발명의 실시를 위한 구체적인 내용

[0015] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 본 발명이 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0016] 도 1은 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템(100)의 구성을 설명하기 위한 블록도이다.

[0017] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템(100)은 사용자(10)의 요구 사항과 데이터의 특성에 따라 지리적으로 분산되어 있는 자원부(30)에 작업을 분배 및 할당한다.

[0018] 도 1을 참고하면, 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템(100)은 데이터 입력/분석부(110), 작업처리 중계부(120), 자원 관리부(130), 스케줄링 관리부(140)를 포함한다.

[0019] 데이터 입력/분석부(110)는 사용자(10)로부터 처리해야 할 데이터와 함께 상기 데이터의 특성과 요구사항을 입력 받고 작업처리 중계부(120)로 작업을 처리할 자원부(30)의 선택을 의뢰하는 역할을 수행한다.

[0020] 작업처리 중계부(120)는 처리할 자원부(30)를 선택하고 전체 작업 수행에 대한 감독 역할을 수행한다. 작업처리 중계부(120)는 데이터 입력/분석부(110)의 요청에 따라 자원 관리부(130)로 자원부(30)의 정보를 요청하고 자원 관리부(130)로부터 자원 정보 결과 메시지를 전달받아 자원부(30)를 선택한 후 스케줄링 관리부(140)로 작업 처리를 의뢰한다.

[0021] 자원 관리부(130)는 이기종 자원을 수집하여 유지하는 역할을 수행한다. 자원 관리부(130)는 자원부(30)로 자원 정보 수집 요청 메시지를 전송하고 자원부(30)로부터 상기 자원 정보 수집 요청 메시지에 대한 자원 정보 응답 메시지를 수신하여 수신된 자원 정보 응답 메시지를 통해 자원부(30)의 정보를 저장 및 갱신, 유지한다. 자원 관리부(130)는 작업처리 중계부(120)로부터 자원 정보 요청이 수신되면 유지하고 있는 자원부(30)의 정보를 나타내는 자원 정보 결과 메시지를 작업처리 중계부(120)로 전송한다.

[0022] 스케줄링 관리부(140)는 작업처리 중계부(120)로부터 작업 처리 의뢰가 수신되면 자원부(30)에게 작업을 분배 및 할당하는 역할을 수행한다. 스케줄링 관리부(140)는 추론을 통해 작업을 자원부(30)로 분배 및 할당한 후 작업을 처리하는 자원부(30)로부터 작업에 대한 처리 결과를 전송 받아 결과를 취합하여 사용자(10)에게 처리 결과를 제공한다.

[0023] 도 2를 참조하여 데이터 입력/분석부(110)의 내부 구성과 동작을 상세하게 설명한다.

[0024] 도 2에 도시한 바와 같이, 데이터 입력/분석부(110)는 사용자(10)로부터 처리하려는 데이터를 입력받는 사용자 데이터 수신기(210)와, 상기 입력된 데이터의 특성을 추출하는 데이터 분석기(230)와, 상기 입력된 데이터에서 자원 사용 예산과 마감기한 정보를 추출하는 요구사항 분석기(250)를 포함한다.

[0025] 사용자 데이터 수신기(210)는 사용자(10)가 처리하려는 데이터를 입력받는다. 여기서, 데이터란 이미지, 문자, 계산 등등 컴퓨팅 자원을 이용하여 처리되어야 할 작업을 총칭한다. 사용자(10)는 처리하고자 하는 데이터를 송신하기에 앞서 데이터의 종류와 크기 등의 데이터 특성, 예산, 마감기한 등의 요구사항을 입력하여 전송한다. 일기예보나 유전자 분석과 같이 자연과학과 공학분야에서의 데이터는 대규모이기 때문에 한번에 전송되어 입력되지 못한다. 이러한 경우 사용자 데이터 수신기(210)는 사용자(10)로부터 데이터를 입력 받은 후 데이터가 자원부(30)에 할당될 때까지 작업 큐(queue)에 저장하고 있다가 자원이 선정되면 작업 큐에 저장된 데이터를 스케줄링 관리부(140)로 전송한다.

[0026] 데이터 분석기(230)는 사용자가 처리하려는 데이터의 특성을 입력 받아 분석한다. 여기서, 데이터의 특성이란 사용자의 데이터가 많은 저장 공간을 요구하는 대용량의 데이터인지, 빠른 계산을 필요로 하는 계산 데이터인지 등에 대한 정보를 포함할 수 있다. 데이터 분석기(230)는 상기 분석된 데이터의 특성을 정수 형태로 변환하고 정수 형태로 변환된 데이터의 특성을 작업처리 중계부(120)로 전송한다.

[0027] 요구사항 분석기(250)는 사용자의 요구사항을 입력 받아 분석한다. 여기서, 사용자의 요구사항이란 사용자가 자원을 이용하여 지출할 수 있는 예산과, 처리하려는 데이터가 언제까지 끝나야 하는지에 대한 마감기한을 포함할 수 있다. 요구사항 분석기(250)는 사용자의 요구사항을 정수 형태로 변환하고 정수 형태로 변환된 요구사항을 작업처리 중계부(120)로 전송한다.

[0028] 도 3을 참조하여 작업처리 중계부(120)의 내부 구성과 동작을 상세하게 설명한다.

[0029] 도 3에 도시한 바와 같이, 작업처리 중계부(120)는 데이터 입력/분석부(110)로부터 자원선택 의뢰를 수신하여 자원 관리부(130)에 현재 사용되지 않는 가용 자원을 요청하는 가용자원 요청기(310)와, 자원 관리부(130)로부

터 가용 자원의 정보 결과 메시지를 받는 가용자원 수신기(330)와, 가용한 자원 목록을 바탕으로 데이터의 특성과 사용자의 요구사항에 적합한 자원을 지식 기반의 추론을 거쳐 선택하는 자원 선택기(350)와, 전반적인 작업 수행에 대한 감독 역할을 수행하는 작업 관리기(370)를 포함 할 수 있다.

[0030] 가용자원 요청기(310)는 자원 관리부(130)에 현재 이용 가능한 자원 목록을 요청한다. 여기서, 가용 자원이란 지리적으로 분산되어 있는 자원부(30) 중 CPU, RAM의 이용률이 일정 비율이상 사용되지 않고 있는 유휴 컴퓨팅 자원을 의미한다.

[0031] 가용자원 수신기(330)는 네트워크를 통하여 자원 관리부(130)로부터 현재 이용 가능한 자원 목록을 수신하고 수신된 가용 자원 목록을 자원 선택기(350)로 전송한다. 가용자원 수신기(330)는 작업 처리를 실패하거나 자원의 갑작스러운 전원 종료 및 네트워크 회선의 단절 등 가용한 상태가 아닐 경우를 대비하여 작업이 처리완료 될 때까지 자원 관리부(130)로부터 가용한 자원 목록을 수신하여 유지한다.

[0032] 자원 선택기(350)는 가용 자원 목록과 데이터 특성, 사용자의 요구사항을 입력 받는다. 여기서, 자원 선택기(350)가 사용자의 요구사항 중 자원을 사용 할 때 지출 할 수 있는 예산(budget)이 넘지 않는 범위 안에서 자원을 선택하는 방법을 여러 가지 규칙으로 정의하여 부합되는 자원을 선택한다. 예를 들어, 데이터의 크기가 대용량이고 충분한 마감기한까지의 시간이 있다면 보조 기억장치는 크고, 빠르지 않더라도 저렴한 자원을 선택하게 된다. 본 발명의 일실시예에서 자원을 선택하는 방법을 표 1과 같은 형태로 규칙을 정의할 수 있으며 시스템 관리자나 사용자에 의해 추가로 규칙을 정의하거나 수정 및 갱신이 용이하게 적용할 수 있다.

표 1

[0033]

Rule	Description
1	DesktopComputer(?x) ^ CPUSpeed(?x, ?y) ^ [?y >= 3000.0] ^ organization(?x, ?z) ^ [?z = usrOrg] → 동일기관고속컴퓨터(?x)
	빠른 처리 속도와 동일한 기관에 속한 자원을 원한다.
2	Server(?x) ^ Bandwidth (?x, ?y) ^ [?y >= 100.0] ^ StorageSize (?x, ?z) ^ [?z > 5000.0] → 대용량고속서버(?x)
	고속 네트워크 환경에서 대용량의 데이터 처리가 가능한 자원을 원한다.
3	Server(?x) ^ CPUSpeed(?x, ?y) ^ [?y >= 3400.0] ^ Cost(?x, ?z) ^ [?z > 15] ^ StorageSize (?x, ?k) ^ [?k < 10] → 작은데이터계산위주서버(?x)
	자원이 비싸더라도 연산에 필요한 데이터 크기는 작고 처리 속도가 빠른 작은 자원을 원한다.
4	Notebook(?x) ^ location(?x, ?y) ^ [?y = usrLoc] ^ Cost(?x, ?z) ^ [?z < 5] → 저가노트북(?x)
	처리해야 할 속도와 크기는 상관없고, 사용자와 동일한 지역에 있는 가격이 저렴한 자원을 원한다.
5	Cluster(?x) ^ numCPU(?x, ?y) ^ [?y >= 5] ^ CPUSpeed(?x, ?z) ^ [z > 3000.0] ^ OSType(?x, ?k) ^ [?k = WinFamily] ^ StorageSize (?x, ?h) ^ [?h >= 1000.0] ^ Bandwidth (?x, ?s) ^ [?s >= 100.0] → 고성능대용량고속클러스터(?x)
	가용 CPU의 개수가 5개 이상이고, OS가 윈도우 계열이며 고속 네트워크 환경의 대용량의 처리가 가능한 자원을 원한다.
	...

[0034] 자원 선택기(350)는 상기 정의된 규칙을 통해 선택된 자원 목록을 스케줄링 관리부(140)로 전송한다.

[0035] 작업 관리기(370)는 사용자의 데이터가 처리되기 위하여 시스템이 시작되어 처리가 완료될 때까지 전체 작업 수행에 대한 감독을 수행한다. 이때, 사용자의 데이터를 처리하는 도중 자원 선택기(350)에서 선택한 자원이 갑작스러운 전원 종료나 네트워크의 단절 등에 의하여 정상적으로 작업을 처리 할 수 없는 경우 작업 관리기(370)는 이를 감지하고 자원 재 선택을 자원 선택기(350)에 명령한다. 또한, 작업 관리기(370)는 작업 처리를 실패할 경우 이에 따른 작업 재전송 및 재처리를 명령할 수 있다.

[0036] 도 4를 참조하여 자원 관리부(130)의 내부 구성과 동작을 상세하게 설명한다.

[0037] 도 4에 도시한 바와 같이, 자원 관리부(130)는 자원의 속성 정보를 요청하는 자원정보 요청기(410)와, 자원의 정적인 속성을 입력 받는 자원정보 수신기(430)와, 작업처리 중계부(120)의 요청에 따라 가용 자원 목록에 대하여 응답하는 자원정보 송신기(450)와, 자원의 정보와 관계를 관리하기 위해 OWL로 표현된 자원 온톨로지 관리기(470)와, 자원에 대한 속성 정보 값을 저장하는 자원정보 저장소(490)를 포함한다.

[0038] 자원정보 요청기(410)는 자원 속성 값에 대하여 주기적으로 자원부(30)에 요청한다. 여기서, 자원의 속성 값이

란 자원의 OS 종류, CPU의 최대 클럭 속도, 보조 기억장치의 크기 등을 포함 할 수 있는 정적인 속성과; CPU 사용률, 주기억 장치 사용률, 보조 기억 장치 가용 공간 등 자원의 현재 상태에 따라 변화 할 수 있는 동적인 속성을 의미한다. 표 2는 자원의 정적인 속성을 나타낸 것이며, 표 3은 동적인 속성을 나타낸 것이다.

표 2

[0039]

Property Name	Description	Value Type	Minimum Value	Unit
ID	자원 식별자	Integer	0	---
OSType	OS의 종류	String	---	---
CPUArchType	CPU의 구조	String	---	---
organization	자원이 소속 된 기관	String	---	---
CPUSpeed	CPU 최대 클럭 속도	Float	700	Megahertz
CPUNumber	CPU 개수	Integer	1	
CacheSize	cache 메모리의 크기	Float	256	Kilo Bytes
RAMSize	주 기억 장치의 크기	Float	32	Mega Bytes
StorageSize	보조기억장치의 크기	Float	10	Giga Bytes
Bandwidth	네트워크의 대역폭	Float	---	Mega BPS
Cost	자원의 사용 금액	Integer	---	---
…	…	…	…	…

표 3

[0040]

Property Name	Description	Value Type	Value Range
CPUStatus	CPU 사용여부	Boolean	true or false
CPUUtilization	CPU 사용률	Float	0~100
RAMUtilization	주기억 장치 사용률	Float	0~100
RAMAvailableSize	주기억 장치 가용 공간	Float	0~Available memory size
StorUtilization	보조 기억 장치 사용률	Float	0~100
StorAvalableSize	보조 기억 장치 가용 공간	Float	0~Available storage size
NetStatus	네트워크의 상태	Float	RTT
JobCompletion Time	Buffer에 쌓여있는 Job이 처리완료 되는 예상시간	Float	0~Estimated Time
Rank	Resource의 순위	Integer	1~10
…	…	…	…

[0041] 본 발명의 일실시예에서 자원 관리부(130)는 자원의 정적인 속성과 동적인 속성에 포함되어 있는 필드를 시스템 관리자에 의해 추가 및 삭제, 수정하여 자원정보를 저장 및 유지할 수 있다. 또한, 자원 관리부(130)는 일정시간이나 지정한 횟수가 지난 후에도 응답이 없는 자원에 대해서는 자동으로 해당 자원 정보를 자원정보 저장소(490)에서 삭제할 수 있다.

[0042] 자원정보 수신기(430)는 자원부(30)로부터 자원의 정적인 속성에 대한 정보 값을 입력 받는다. 여기서, 자원의 정적인 속성은 자원정보 요청기(410)에 의한 주기적인 요청 메시지에 대한 응답일 수도 있고 자원부(30)가 자신의 정보가 변경되었음을 감지하여 이벤트를 발생시켜 자동으로 송신될 수도 있다. 그리고, 자원정보 수신기(430)는 수신된 자원 정보를 자원정보 저장소(490)에 저장한다.

[0043] 자원정보 송신기(450)는 작업처리 중계부(120)의 가용자원 요청기(310)로부터 가용 자원 요청 메시지를 수신하면 현재 이용 가능한 자원을 자원정보 저장소(490)에서 검색하여 조건에 부합되는 자원을 선별해서 자원 목록 정보를 생성한 후 작업처리 중계부(120)의 가용자원 수신기(330)로 전송한다.

[0044] 자원 온톨로지 관리기(470)는 자원의 종류에 따라 자원을 구분하게 되는데 W3C에서 제정한 온톨로지 표준 언어인 OWL을 이용하여 자원을 구분할 수 있다.

[0045] 도 5는 자원을 클래스 형태로 도시한 것으로, 도시한 바와 같이 최상위 클래스는 Resource이며, 자원의 종류에

따라 계산 위주의 데이터를 처리하기 위한 ComputingResource와, 대용량의 데이터를 처리하기 위한 DataResource로 구분할 수 있다. 그리고, 새로운 종류의 자원이 추가될 경우 새로운 타입의 자원을 ResourceType#N 클래스로 정의해서 온톨로지에 포함시켜 시스템 확장이 용이하고 유연한 구조를 가질 수 있다.

[0046] 자원정보 저장소(490)는 자원에 대한 속성 정보 값을 도 5와 같이 구축된 자원 온톨로지 형태 중 하나로 표 2와 표 3에 도시된 자원의 정적 속성과 동적인 속성으로 저장한다.

[0047] 도 6을 참조하여 스케줄링 관리부(140)의 내부 구성과 동작을 상세하게 설명한다.

[0048] 도 6에 도시한 바와 같이, 스케줄링 관리부(140)는 자원부(30)로부터 자원의 상태 값을 받는 자원상태정보 수신기(610)와, 수신된 자원의 상태 값을 토대로 자원부(30)를 선별하는 추론 엔진부(630)와, 작업 처리 중계부(120)의 자원 선택기(350)로부터 선택된 자원을 작업을 처리하기 위한 자원으로 구성하고 추론 엔진부(630)에서 선별된 자원에게 작업을 할당 및 분배하는 작업 분배기(650)와, 자원부(30)로부터 완료된 작업을 취합하여 파일로 만드는 처리완료 데이터 취합기(670)를 포함한다. 여기서, 추론 엔진부(630)는 시스템 관리자나 전문가에 의해 규칙을 추가 및 삭제, 수정할 수 있는 규칙 관리기(631)와, 정의된 규칙 수행에 필요한 조건을 만족하는 사실을 탐색하고 새로운 사실을 도출하는 규칙 해석기(633)와, 정제된 자원들 중에서 질의를 처리하는 질의 처리기(635)를 포함한다.

[0049] 자원상태정보 수신기(610)는 자원부(30)로부터 자원의 동적인 상태 정보 값을 입력 받는다. 여기서, 자원의 동적인 상태 정보 값이란 표 3에 도시되어 있는 동적인 자원 상태로서 CPU의 사용 여부, CPU 사용률, 주기억 장치 사용률, 보조 기억 장치 사용률, 네트워크의 상태 등을 포함할 수 있다. 자원상태정보 수신기(610)는 수신된 자원상태정보를 데이터 세트 형태로 추론 엔진부(630)로 전송한다.

[0050] 추론 엔진부(630)의 규칙 관리기(631)는 수행되어야 할 규칙을 시스템 관리자나 전문가에 의해서 추가로 정의하거나 수정 및 삭제할 수 있다. 여기서, 규칙은 표 2와 표 3에 도시되어 있는 자원의 정적인 속성과 동적인 속성을 조합하여 도 5에 도시되어 있는 자원 클래스 중 하나로 개체(individual)를 만들 수 있다. 이 때, 시스템 관리자나 전문가가 자원의 동적인 상태와 네트워크 상황과 같은 주변 환경에 따라 규칙을 새로 추가하거나 삭제 또는 수행되어야 할 우선 순위를 변경하여 적용시킬 수 있다. 본 발명의 일실시예에 따른 동적 작업 스케줄링을 위한 규칙은 표 4와 같다.

표 4

[0051]

Rule	Domain rule
1	Resource(?x) ∧ StorAvalableSize(?x, ?y) ∧ [?y > jobSize] → AvailableResource(?x)
	남아있는 디스크 공간이 처리 할 job의 크기보다 크면 가용 자원으로 정의한다.
2	Resource(?x) ∧ CPUSpeed(?x, ?y) ∧ [?y > 3.0] ∧ CacheSize(?x, ?z) ∧ (?z > 4096) ∧ CPUUtilizationr(?x, ?k) ∧ [?k < 30] → HighSpeedLowWorkloadResource(?x)
	CPU의 최대 클럭 속도가 3.0GHz 이상이고, cache의 크기가 4096 kilobytes 이상이고, CPU 사용률이 30% 미만이면 낮은부하의고속처리CPU자원으로 정의한다.
3	Resource(?x) ∧ CPUUtilization(?x, ?y) ∧ [?y > 90] ∧ StorUtilization(?x, ?z) ∧ [?z > 90] ∧ RAMUtilization(?x, ?k) ∧ [?k > 90] ∧ NetStatus(?x, ?n) ∧ [?n > threshold] → hasRank(?x, 1)
	CPU 사용률, 디스크 사용률, 메모리 사용률이 모두 90%를 초과하고 네트워크가 혼잡 상태이면 자원의 Rank는 1이다.
4	Resource(?x) ∧ CPUUtilization(?x, ?y) ∧ [?y > 40] ∧ [?y < 50] ∧ StorUtilization(?x, ?z) ∧ [?z > 40] ∧ [?z < 50] ∧RAMUtilization(?x, ?k) ∧ [?k > 40] ∧ [?k < 50] ∧ NetStatus(?x, ?n) ∧ [?n < threshold] → hasRank(?x, 5)
	CPU 사용률, 디스크 사용률, 메모리 사용률이 40~50% 이고, 네트워크가 정상 상태이면 자원의 Rank는 5이다.
5	Resource(?x) ∧ CPUUtilization(?x, ?y) ∧ [?y < 10] ∧ StorUtilization(?x, ?z) ∧ [?z < 10] ∧ RAMUtilization(?x, ?k) ∧ [?k < 10] ∧ NetStatus(?x, ?n) ∧ [?n < threshold] → hasRank(?x, 10)
	CPU 사용률, 디스크 사용률, 메모리 사용률이 모두 10% 미만이고 네트워크가 정상 상태이면 자원의 Rank는 10이다.

6	Resource(?x) ∧ NetStatus(?x, ?y) ∧ [?y < threshold] ∧ bandwidth(?x, ?z) ∧ [?z >= 100] → HighSpeedNetworkResource(?x)
	네트워크 환경이 100Mbps 이상을 지원하고, 임계값 보다 작은 RTT를 가지면 고속네트워크자원으로 정의한다.
…	…

[0052] 추론 엔진부(630)의 규칙 해석기(633)는 표 4에 도시되어 있는 규칙 수행에 필요한 조건을 만족하는 사실을 탐색하고 새로운 사실을 도출하는 역할을 수행한다.

[0053] 추론 엔진부(630)의 질의 처리기(635)는 규칙 해석기(633)에 의해 도출된 새로운 사실을 바탕으로 작업 처리의 효율성을 극대화하기 위하여 질의를 수행해서 현재의 작업을 분배할 자원을 선별한다. 그리고, 질의 처리기(635)는 선별된 자원에 대한 인덱스 정보를 작업 분배기(650)로 전송하는 역할을 수행한다.

[0054] 작업 분배기(650)는 작업처리 중계부(120)의 자원 선택기(350)로부터 선별된 자원과 질의 처리기(635)에 의해 선별된 자원을 입력받는다. 여기서, 자원 선택기(350)로부터 선별된 자원은 사용자의 작업 처리 의뢰 메시지를 수신한 순간에 전송 받으며 작업 처리를 위한 분산 자원으로 구성한다. 만약, 작업처리 중계부(120)의 작업 관리기(370)의 명령에 의해 재 선택된 자원 목록을 수신하면 필요에 따라 수신된 자원으로 재구성 할 수 있다. 이 때, 추론 엔진부(630)를 통하여 질의 처리기(635)에서 최종적으로 선별된 자원부(30)에게 작업을 분배 및 할당한다.

[0055] 처리완료 데이터 취합기(670)는 자원부(30)로부터 처리가 완료된 데이터를 취합하여 파일로 만들어서 사용자에게 반환하는 역할을 수행한다.

[0056] 도 7을 참조하여 자원부(30)의 내부 구성과 동작을 상세하게 설명한다.

[0057] 도 7에 도시한 바와 같이, 자원부(30)는 스케줄링 된 작업을 입력 받아 처리하는 응용 처리기(710)와, 자원의 정적/동적인 상태를 측정하는 자원상태 측정기(730)와, 측정된 자원의 상태 정보 값을 전송하는 자원상태 전송기(750)를 포함한다.

[0058] 응용 처리기(710)는 스케줄링 관리부(140)의 작업 분배기(650)로부터 분배 및 할당된 사용자의 작업을 입력 받아 처리하는 역할을 수행하고 처리 완료된 작업을 처리완료 데이터 취합기(670)로 전송한다.

[0059] 자원상태 측정기(730)는 자원 관리부(130)의 자원정보 요청기(410)에 의해 요청된 자원 정보 요청 메시지를 수신하면 표 2와 표 3에 도시된 바와 같이 자원의 정적/동적인 속성을 측정한다.

[0060] 자원상태 전송기(750)는 측정된 자원의 정적/동적인 속성 정보 값을 자원 관리부(130)의 자원정보 수신기(430)와 스케줄링 관리부(140)의 자원상태정보 수신기(610)로 전송한다.

[0061] 상기한 구성에 의한 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 이기종의 자원으로 이루어진 컴퓨팅 환경에서 사용자의 작업을 처리하기 위하여 지리적으로 분산되어 있는 자원들의 정보와 관계를 온톨로지로 구축하고, 다양한 규칙으로 정의된 스케줄링을 통한 데이터 분배 및 할당에 의하여 사용자의 요구사항과 데이터 특성에 적합한 자원을 선택 할 수 있다.

[0062] 도 8은 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 방법의 과정을 설명하기 위한 도면이다.

[0063] 도 8을 참조하여 본 발명의 일실시예에 있어 스케줄링 할 자원을 선별하는 과정을 상세하게 설명하면 다음과 같다. 여기서, 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 방법은 도 1에 도시한 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템, 특히 도 6에 도시한 스케줄링 관리부(140)에 의해 실행될 수 있다.

[0064] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 데이터 입력/분석부(110)의 데이터 분석기(230)로부터 분석된 데이터의 특성인 데이터 크기를 입력 받는다(S801). 그리고, 자원 온톨로지 OWL 파일을 적재하고(S802) 추론이 가능한 형태로 데이터의 정적 속성과 동적 속성을 바탕으로 자원 개체를 생성한다(S803).

[0065] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 자원을 선별하기 위하여 표 4에 정의된 규칙을 적용하는데, 여기에서는 입력 받은 데이터의 크기가 자원의 동적인 속성 중 보조 기억 장치 가용 공간 보다 작은지 판단하고(S804) 상기 단계(S804)의 조건을 만족하는 자원들을 가용 자원 목록으로 추가한다

(S805).

[0066] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 자원의 동적인 속성을 바탕으로 작업 완료 예상시간, CPU 이용률, 주 기억 장치 이용률, 보조 기억 장치의 이용률을 차례로 계산한다(S806~S809). 상기한 과정(S806~S809)에서 계산된 자원의 동적인 속성을 이용하여 표 4에 정의 된 추론 규칙을 적용하여 가용 자원 목록에 해당하는 자원의 등급을 결정한 후(S810), 자원의 등급을 바탕으로 높은 순으로 가용 자원 목록을 정렬한다(S811).

[0067] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 상기 단계(S811)에서 정렬된 가용 자원 목록에서 같은 등급의 자원이 존재할 경우 작업 완료 예상시간이 작은 자원에 우선 순위를 두고(S812), 같은 등급의 자원 중 작업 완료 예상시간이 같은 자원이 존재할 경우 네트워크의 상태가 좋은 자원에 우선 순위를 둔다(S813).

[0068] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 가용 자원 목록 중 순위가 가장 높은 자원을 최종적으로 선별하여 최종 선별 된 자원 정보 인덱스를 스케줄링 관리부(140)의 작업 분배기(650)로 전송한다(S814).

[0069] 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템은 상기한 과정에 의한 스케줄링 방법 이외에도 시스템 관리자나 전문가에 의해 표 4에 도시한 바와 같이 추가로 규칙을 정의하거나 수정 및 갱신, 우선 순위를 변경하여 적용함으로써 동적인 자원의 상태 변화에 따라 다양한 규칙을 적용할 수 있다.

[0070] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

[0071] 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면의 간단한 설명

[0072] 도 1은 본 발명의 일실시예에 따른 시멘틱 컴퓨팅 기반의 동적 작업 스케줄링 시스템의 구성을 도시한 도면이다.

[0073] 도 2는 도 1에 도시한 데이터 입력/분석부(110)의 일례를 도시한 도면이다.

[0074] 도 3은 도 1에 도시한 작업처리 중계부(120)의 일례를 도시한 도면이다.

[0075] 도 4는 도 1에 도시한 자원 관리부(130)의 일례를 도시한 도면이다.

[0076] 도 5는 추론에 필요한 개체를 생성하기 위한 자원 온톨로지 모형의 일례를 도시한 도면이다.

[0077] 도 6은 도 1에 도시한 스케줄링 관리부(140)의 일례를 도시한 도면이다.

[0078] 도 7은 도 1에 도시한 자원부(30)의 일례를 도시한 도면이다.

[0079] 도 8은 정의된 규칙을 바탕으로 자원의 속성 정보를 이용한 동적 스케줄링 과정의 일례를 설명하기 위한 도면이다.

[0080] <도면의 주요 부분에 대한 부호의 설명>

[0081] 110: 데이터 입력/분석부

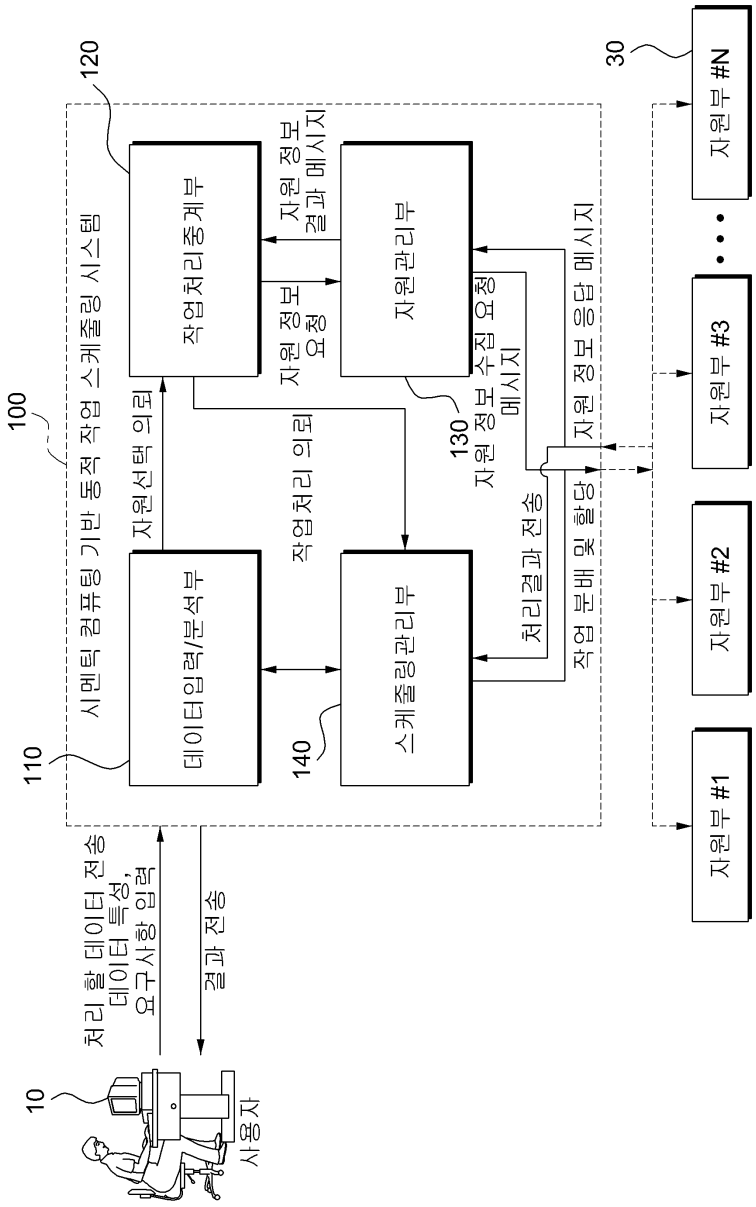
[0082] 120: 작업처리 중계부

[0083] 130: 자원 관리부

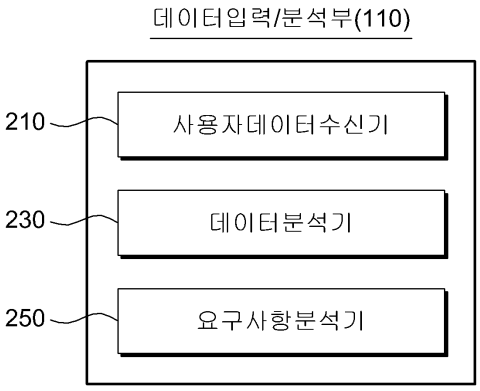
[0084] 140: 스케줄링 관리부

도면

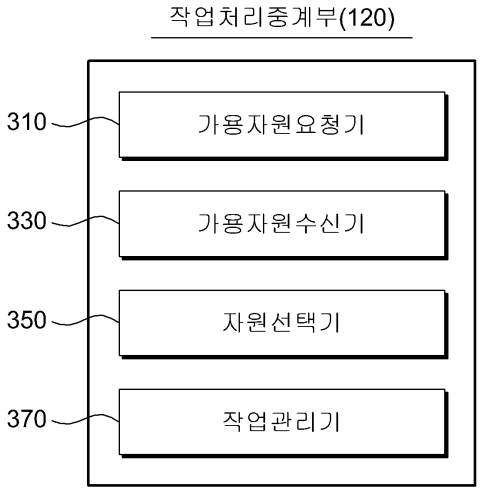
도면1


100
시멘틱 컴퓨팅 기반 동적 작업 스케줄링 시스템
110
데이터입력/분석부
120
작업처리중계부
130
자원관리부
140
스케줄링관리부
자원선택 의뢰
작업처리 의뢰
자원 정보 요청
자원 정보 결과 메시지
자원 정보 수집 요청 메시지
처리결과 전송
작업 분배 및 할당
자원 정보 응답 메시지
10
사용자
처리 할 데이터 전송
데이터 특성,
요구사항 입력
결과 전송
30
자원부 #1
자원부 #2
자원부 #3
자원부 #N

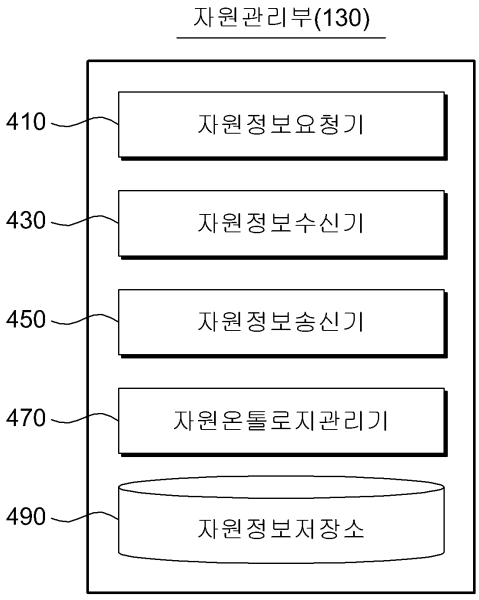
도면2


데이터입력/분석부(110)
210
사용자데이터수신기
230
데이터분석기
250
요구사항분석기

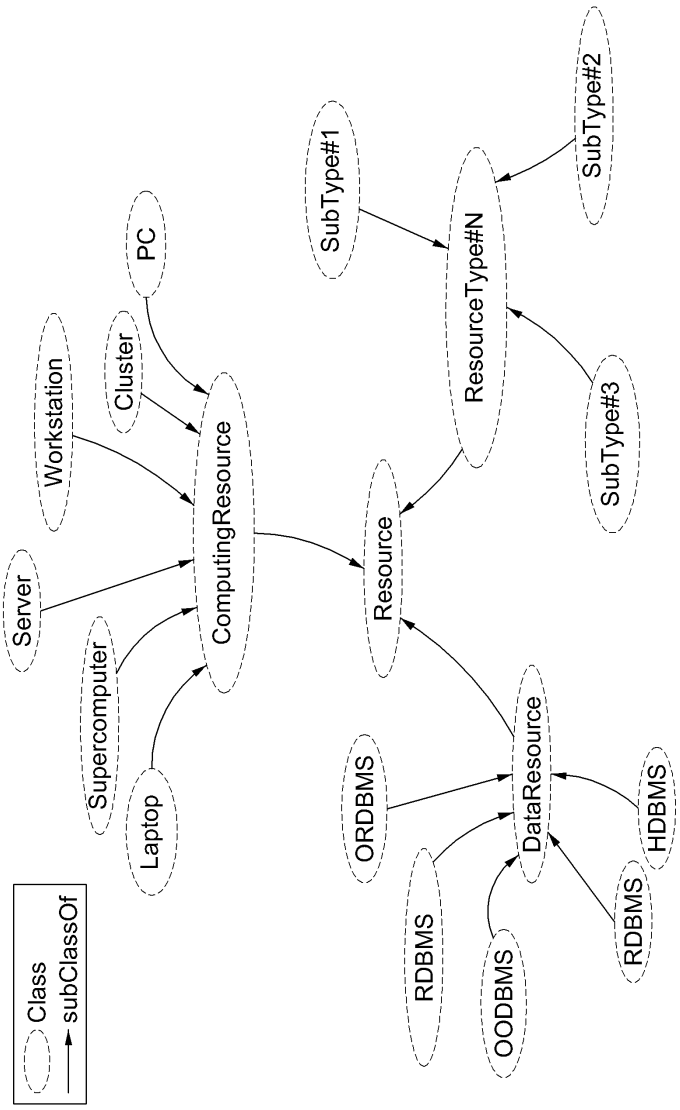
도면3


작업처리중계부(120)
310
가용자원요청기
330
가용자원수신기
350
자원선택기
370
작업관리기

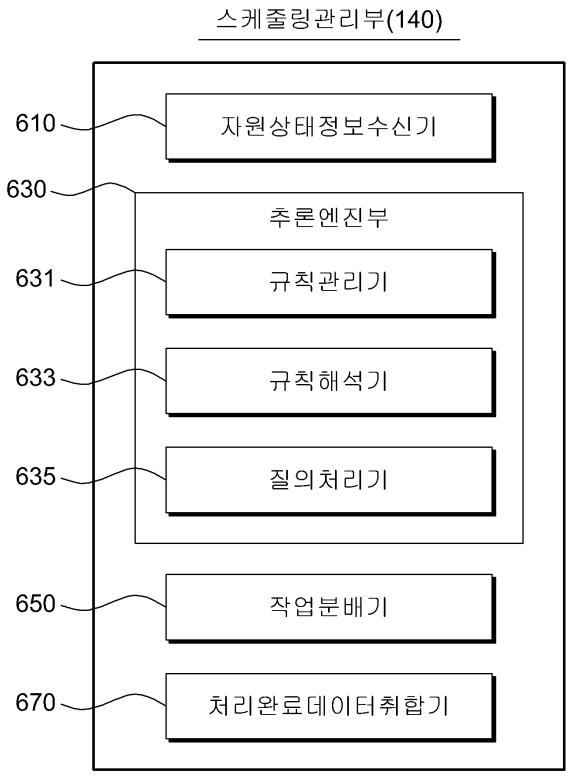
도면4


자원관리부(130)
410
자원정보요청기
430
자원정보수신기
450
자원정보송신기
470
자원온톨로지관리기
490
자원정보저장소

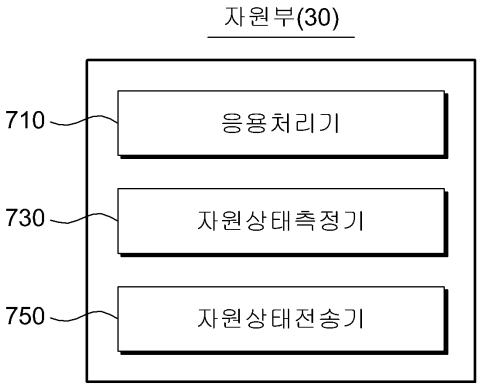
도면5


Class
subClassOf
Server
Workstation
Cluster
PC
Supercomputer
Laptop
ComputingResource
Resource
SubType#1
SubType#2
ResourceType#N
SubType#3
ORDBMS
RDBMS
OODBMS
DataResource
HDBMS
RDBMS

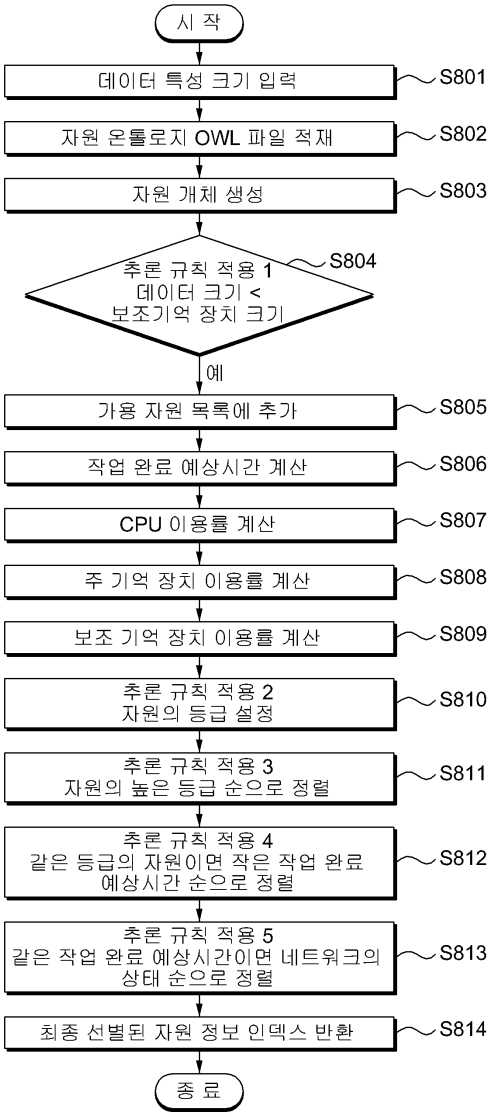
도면6


스케줄링관리부(140)
610
자원상태정보수신기
630
추론엔진부
631
규칙관리기
633
규칙해석기
635
질의처리기
650
작업분배기
670
처리완료데이터취합기

도면7


자원부(30)
710
응용처리기
730
자원상태측정기
750
자원상태전송기

도면8


시 작
데이터 특성 크기 입력
S801
자원 온톨로지 OWL 파일 적재
S802
자원 개체 생성
S803
추론 규칙 적용 1
데이터 크기 <
보조기억 장치 크기
S804
예
가용 자원 목록에 추가
S805
작업 완료 예상시간 계산
S806
CPU 이용률 계산
S807
주 기억 장치 이용률 계산
S808
보조 기억 장치 이용률 계산
S809
추론 규칙 적용 2
자원의 등급 설정
S810
추론 규칙 적용 3
자원의 높은 등급 순으로 정렬
S811
추론 규칙 적용 4
같은 등급의 자원이면 작은 작업 완료
예상시간 순으로 정렬
S812
추론 규칙 적용 5
같은 작업 완료 예상시간이면 네트워크의
상태 순으로 정렬
S813
최종 선별된 자원 정보 인덱스 반환
S814
종 료