

# 대사 증후군 질병 진단을 위한 정책 기반 그리드 워크플로우 관리 시스템

정우람\*, 김회영\*, 윤찬현\*, 최연식\*\*, 홍혁기\*\*

\*카이스트 정보통신과

\*\*한국 전자 부품 연구원

e-mail : {wrjung, hoeyoung, chyoun}@icu.ac.kr

## Policy-based Grid Workflow Management System for Metabolic Syndrome Identification

Woo-Ram Jung\*, Hoe-Youn Kim\*, Chan-Hyun Youn\*, Yeon-Sik Choi\*\*, Hyuck-Ki Hong\*\*

\*Dept. of Information and Communication, KAIST

\*\*Dept of Medical IT Convergence Research Center, KETI

### 요 약

컴퓨팅 및 네트워크 기술이 발전함에 따라 그리드 응용의 복잡도 및 컴퓨팅 요구사항이 커지게 되었으며, 이로 인해 그리드 워크플로우 관리 기술의 중요성이 더욱 부각되었다. 이러한 복잡한 그리드 응용을 위해 다양한 그리드 워크플로우 관리 시스템이 개발되었으나 사용자의 요구사항에 대한 고려, 안정성 그리고 사용자와의 상호작용 등의 기능이 부족하였으며 특히 자원관리 기능에 대한 지원이 부족하였다. 본 논문에서는 기존의 응용 친화적인 워크플로우 관리 시스템과 그리드 워크플로우 관리 시스템을 정책 기반 메커니즘을 이용, 통합하여 새로운 Hybrid 형태의 워크플로우 관리 시스템을 제안하였으며, 제안하는 방법이 어떻게 사용자의 요구수준을 만족시킬 수 있는지를 서술하고 이를 대사증후군을 위한 질병진단 시스템에 적용하여 실험적으로 증명하였다.

### 1. 서론

그리드 컴퓨팅[1] 기술은 이 기종의 분산된 네트워크 환경에서 데이터를 공유하고 처리하는 기술로써 기존의 컴퓨팅 기술로는 풀기 어려운 다양한 응용의 개발을 가능하게 하였다. 그리드 워크플로우[7] 기술은 이와 같은 그리드 상의 복잡한 응용을 작은 작업들로 나누고 이를 스케줄링 기술을 통해 정해진 흐름에 맞게 효과적으로 수행하도록 하는 기술을 말한다. 하지만 지금까지의 그리드 워크플로우 기술은 컴퓨팅 성능에만 초점을 맞추었기 때문에, 기존의 응용레벨에서의 워크플로우에서 제공하는 유저 인터페이스나 작업의 흐름을 추상적으로 표현하는 기능이 부족하였으며 수행 중 사용자와의 상호작용이 어렵다는 단점이 있었다. 뿐만 아니라 시간적, 공간적으로 불안정적이며 동적인 네트워크 안에서 워크플로우 시스템이 자원관리, 스케줄링 시 사용자의 서비스 수준을 고려한 기능적인 요소가 부족하기 때문에 이를 보완하는 시스템 연구가 요구되었다. 본 논문에서는 이러한 문제점을 보완하고 사용자 응용 서비스의 요구 수준을 만족할 수 있도록 기존의 두 가지 워크플로우 시스템인 Kepler[9] 및 Pegasus[3][4][8] 워크플로우 관리 시스

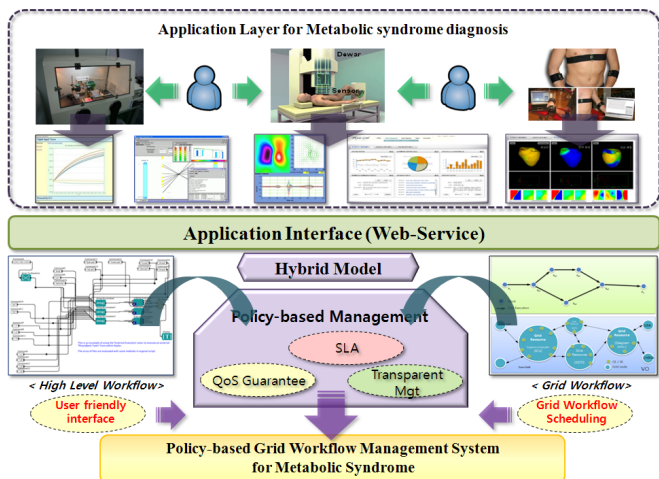
템을 이용해 자원관리 시스템의 제어를 효과적으로 할 수 있도록 새로운 Hybrid 구조의 워크플로우 관리 시스템을 제안하고 이를 시험적으로 구현하여 성능을 평가하였다.

정책기반 Hybrid 형 워크플로우 관리 시스템이란 기존에 개발된 응용 레벨의 워크플로우 관리 시스템의 장점과 그리드 워크플로우 관리시스템의 장점을 수용해 통합된 형태의 워크플로우 관리 시스템을 정책 기반 관리 메커니즘을 이용하여 제공하는 것을 말한다. 응용레벨 워크플로우의 사용자 친화적인 특징과 그리드 워크플로우의 스케줄링 기능을 통해 사용자의 QoS 를 보장하고 적절한 정책 선정을 통해 동적으로 변화하는 환경에서의 사용자 요구수준을 맞추어 준다. 본 논문에서는 이러한 워크플로우 응용의 실행 시 사용자의 요구사항에 따라 적절한 정책을 선정하기 위한 알고리즘을 제안하였고, 이를 실제 구현 및 적용하였다. 우리는 제안된 시스템의 성능 평가를 위해 대사증후군 질병 진단 시스템의 분석 모듈을 워크플로우 형태로 구성하여 실험에 활용하였다. 대사 증후군은 비만, 당뇨병, 고혈압, 고지혈증 등이 한 개인에서 동시에 나타나게 되는 질병을 말하는 것으로 매

년 질병 발생 빈도가 급격하게 늘어나고 있다. 특히 대사증후군의 하나인 당뇨병과 허혈성 심질환에 의한 사망률이 급격하게 증가하고 있는 추세이다. 본 워크플로우 관리 시스템은 이러한 대사 질환 진단을 위한 응용 관리 및 분석 시스템으로 결과 데이터를 이용하여 의사가 더 정확한 진단을 할 수 있도록 돕는다. 본 논문에서는 이러한 질병 진단 응용에 대한 실험을 통해 제안된 시스템이 사용자의 요구 수준에 맞추어 적절한 서비스를 제공할 수 있는 것을 입증하였다.

**2. 대사 증후군 질병 진단을 위한 워크플로우 관리 시스템 모델 및 구조**

대사 증후군 질병 진단을 위한 정책 기반 Hybrid 워크플로우 관리 시스템은 기존의 응용 레벨의 워크플로우의 장점인 사용자 친화적인 면, 즉 사용자와 동적으로 상호작용하기 용이하다는 점과 기존 그리드 워크플로우 관리 시스템의 고성능 스케줄링 기능을 통합하여 유저의 다양한 요구 수준을 만족 시킬 수 있는 새로운 워크플로우 관리 시스템을 말한다. 그림 1 은 제안하는 관리 시스템의 개념 모델을 보인다. 하부는 워크플로우 관리 시스템의 개념 모델로써 Kepler 와 Pegasus 워크플로우 관리 시스템 및 이를 통합하는 정책기반 관리 모듈을 포함한다. 이는 시스템의 유연한 통합을 위해 정책관리 모듈을 두어 특정 응용에 대한 유저의 요구사항을 반영할 수 있도록 정책을 정의하고 이를 적용 할 수 있도록 하기 위함이다.

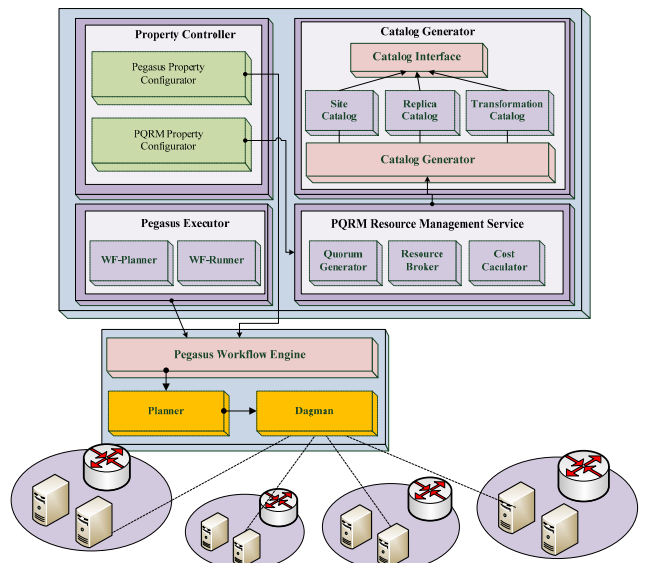


(그림 1) 시스템 모델

상부의 응용 레이어는 대사 증후군 질병 진단을 위한 다양한 질병 분석 서비스들과 세포 열측정 센서를 보여준다. 열 측정 센서는 환자의 세포에 대한 열측정 정보 및 CO<sub>2</sub>, pH 등의 데이터를 질병 진단 시스템에 실시간으로 전송할 수 있으며, 이러한 데이터는

워크플로우 관리 시스템에 의해 분석되어 의사에게 웹을 통해 결과 정보를 가시화 한다.

Hybrid 그리드 워크플로우 관리 시스템은 크게 세 부분으로 나누어 볼 수 있다. 우선 Kepler 시스템을 위한 인터페이스로 이는 Kepler 에서 그리드 워크플로우 관리 모듈을 호출 및 수행 할 수 있도록 해주는 Pegasus Actor 와 Director 가 있다. Pegasus Actor 는 그리드의 자원을 표현하는 하나의 워크플로우 노드를 의미하며 Director 는 이를 수행하는 방법을 정의한 노드이다. 두 번째로 Pegasus 와의 연동을 위한 자원 카탈로그 생성 모듈과 페가수스 프로퍼티 제어 모듈이 있다. 이는 선정된 정책을 기반으로 적절한 워크플로우 스케줄링 방법을 정의 하고 PQRMP<sup>PLUS</sup>[5] 에서 선정된 리소스의 정보를 Pegasus 에 넘겨주는 역할을 한다. 마지막으로 유저의 요구사항에 기반하여 정책을 결정하고 적용하는 정책 관리 모듈 및 정책 저장소가 존재한다. 정책은 웹 기반 인터페이스를 통해 관리자의 권한을 통해 언제든지 새롭게 정의 되거나 지워질 수 있다.



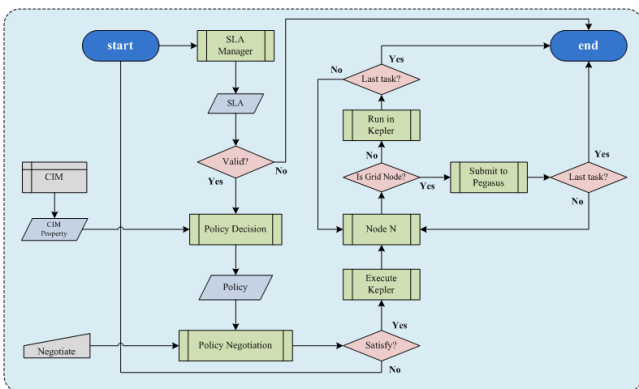
(그림 2) 워크플로우 관리 컴포넌트

그림 2 는 전체 구조에서 Pegasus 와 PQRMP<sup>PLUS</sup> 의 통합 모델의 구현 컴포넌트를 보인다. PQRMP<sup>PLUS</sup> 자원 관리 서비스는 사용자의 요구사항이 반영되어 Pegasus 에게 물리적인 자원에 대한 정보를 제공한다. 이와 같은 정보는 Site Catalog 생성자에 의해 Pegasus 가 해석할 수 있는 형태로 변환된다. Replica Catalog 와 Transformation Catalog 는 Pegasus 를 통해 워크플로우 응용을 수행하기 위한 정보 즉, 각 노드를 통해 수행될 응용의 위치와 입출력 데이터 정보를 생성한다. Property Controller 는 사용자 수준에 따라 적절한 스케줄링 기법을 설정하고 PQRMP<sup>PLUS</sup> 의 자원 선정

방법을 결정하는 역할을 한다. Catalog 및 Property 정보가 모두 준비되면 워크플로우 응용은 Pegasus Executor 에 의해 그리드 상에서 수행된다. 그리드 상의 각 노드에 존재하는 Pegasus Worker 는 스케줄링된 정보를 기반으로 워크플로우 응용을 수행하게 되며, 응용이 수행되어 완료되기까지의 일련의 과정은 로그파일을 통해 모니터링이 가능하다.

**3. 정책결정 과정 및 워크플로우의 수행**

위에서 설명한 워크플로우 수행은 정책 관리자에 의해 미리 정의된 정책을 PQRMP<sup>PLUS</sup> 미들웨어에 적용함으로써 수행되게 된다. 이는 사용자의 요구사항을 반영할 수 있도록 적절한 정책을 결정하는 절차가 선행되어야 함을 의미한다. 그림 3 은 이와 같은 정책이 선정되는 과정 및 전체 워크플로우 응용의 수행 과정을 보인다. 우선 SLA Manager 는 유저의 요구수준을 반영할 수 있도록 이를 해석해주는 역할을 한다. 유저의 요구수준이 결정되면 Policy Decision 모듈이 요구사항에 가장 적절한 정책을 선정하게 된다. 결정된 정책과 그에 따른 비용에 대해 사용자가 동의할 경우 워크플로우를 수행할 수 있게 된다. 전체 워크플로우 응용은 Kepler 위에서 수행을 시작한다. Pegasus Director 에 의해 Kepler 상에서 수행되는 노드 중 그리드 워크플로우 관리 시스템의 수행을 요구하는 노드(즉, 고성능 컴퓨팅 혹은 데이터 저장소가 필요한 노드)가 발견되면 Kepler 시스템은 PQRMP<sup>PLUS</sup> 의 통합 모듈 서비스의 호출을 통해 Pegasus 에 그리드 응용을 제출 및 수행하게 된다.

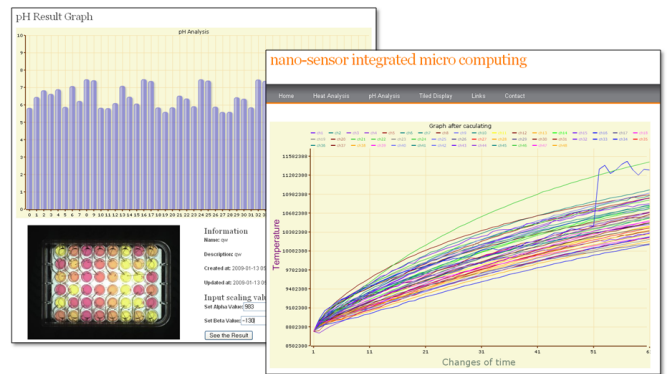


(그림 3) 정책 결정 과정 및 워크플로우 수행 절차

이와 같이 계층적인 구조로 이루어진 워크플로우 관리 시스템은 위에서 언급한 바와 같이 데이터 추출, 유저와의 상호작용 등 사용자 친화적인 면과 복잡한 그리드 응용을 수행하는데 있어서 기존의 시스템에 비교하여 수행 시간 및 사용의 편의성에 있어서 좋은 성능을 보인다.

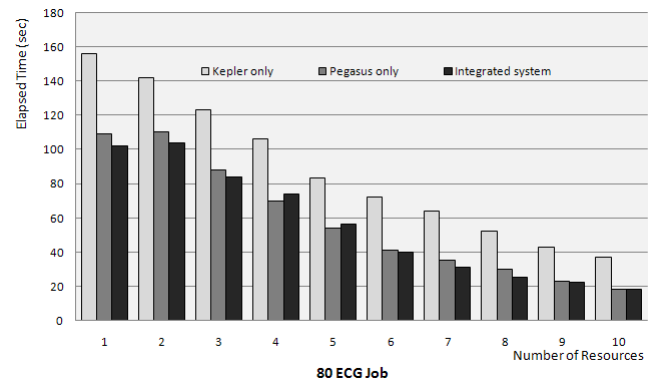
**4. 프로토타입 구현 및 성능 평가**

우리는 제안하는 시스템의 성능 평가 및 그리드 응용에의 적용을 위해 대사 증후군 질병 진단 E-health 응용을 위한 워크플로우를 구성하고 이를 통해 성능 평가를 하였다. 이는 환자로 부터 실시간으로 취득된 세포 열 데이터 분석하여 당뇨병, 심장 질환 등의 질병을 진단하는데 도움을 준다. 그림 4 는 제안하는 워크플로우 시스템을 통해 분석된 결과 데이터를 웹 기반 인터페이스를 이용해 나타낸 결과이다.



(그림 4) 세포열 측정 데이터 분석 결과 화면

그림은 열 측정 장치로부터 얻는 고해상도 이미지 데이터와 열 측정 데이터를 워크플로우 시스템을 통해 분석한 결과 화면으로 왼쪽 화면은 이미지 분석을 통해 pH 분포를 나타내주며, 오른쪽 화면은 실시간으로 생성되는 데이터를 분석하여 이를 가시화 한 모습을 나타낸다. 그림 5 는 자원의 수에 대한 워크플로우 수행 시간을 기존 워크플로우 시스템과 비교 분석한 것이다. 결과에서 볼 수 있듯이 제안하는 시스템은 Kepler 만을 사용한 시스템보다는 상당히 우수한 성능을 제공하였고, Pegasus 만을 사용한 시스템과 거의 동등한 성능을 보였다.



(그림 5) 기존 워크플로우 시스템과의 성능 비교

Pegasus 만을 이용한 시스템은 고성능 UI 를 제공하지 못하는 단점이 있으나, 그리드 네트워크에서

최적화된 상태로 워크플로우가 수행되므로 비록 정책이 적용되지는 않았지만 성능면에서는 좋은 결과를 얻었다. 이는 Kepler 에서의 수행 시간 및 정책 기반 관리 기법에 대한 추가 처리시간이 약간의 성능 저하를 야기시켰다고 볼 수 있으며 유저 친화적 UI 를 사용하는 것에 대한 trade-off 라고 할 수 있겠다.

## 5. 결론

본 논문에서는 기존의 그리드 워크플로우 시스템에 대한 한계를 극복하기 위한 새로운 워크플로우 관리 기법을 제안하였다. 이는 기 개발된 응용 친화적인 워크플로우 관리 시스템과 그리드 워크플로우 관리 시스템을 정책기반 관리 기법을 이용해 통합한 Hybrid 형태의 관리 기술이다. 사용자의 요구 수준에 따라 적절한 정책을 선택하여 자원 선정 및 워크플로우를 스케줄링 하여 응용의 수행 성능을 보장하며, 사용자 친화적인 인터페이스를 제공할 수 있다. 우리는 제안하는 시스템의 프로토타입을 개발하여 이를 대사 증후군 질병 진단 E-health 응용에 적용하였으며, 이를 위한 워크플로우 응용을 설계 및 구현하였다. 마지막으로 구현된 워크플로우의 수행시간을 기존의 워크플로우 관리 시스템과 비교하여 제안하는 시스템이 기존의 시스템에 비해 기능적, 성능적으로 뛰어난 것을 보였다.

## 6. 감사의 말

본 연구는 지식경제부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음 [IITA-2009-C1090-0902-0014, 2006-S-075-03].

## 참고문헌

- [1] Foster, I. and Kesselman, C. (eds.). "The Grid: Blueprint for a New Computing Infrastructure". Morgan Kaufmann, 1999.
- [2] Web-based system for Advanced Heart Disease Identification using Grid Computing Technology, Chang Hee Han, Chan-Hyun Youn, Wooram Jung, IEEE CBMS 2008
- [3] E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, S. Patil, M. H. Su, K. Vahi, M. Livny. Pegasus: Mapping Scientific Workflow onto the Grid. Across Grids Conference 2004, Nicosia, Cyprus, 2004.
- [4] E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, K. Vahi. Mapping Abstract Complex Workflows onto Grid Environments. Journal of Grid Computing, 1:25-39, Kluwer Academic Publishers, Netherlands, 2003.
- [5] Dong Su Nam et al. QoS-constraint Quorum Management scheme for a Grid-based Electrocardiogram, LNCS 3043 pp.352-359, 2004.
- [6] Jay Jie Dong. Grid Workflow-Integrated Resource Management System with a Policy Negotiator. 2008.

[7] A Taxonomy of Scientific Workflow Systems for Grid Computing J Yu, R Buyya 1 In Workflow in Grid Systems Workshop, GGF-10, Berlin, March 9, 2004

[8] GriPhyN. <http://www.griphyn.org>

[9] <http://www.kepler-project.org>