

그리드 환경에서 서비스 기반 워크플로우 시스템 설계 및 구현

최주호^o 권원원 류소현 정창성
고려대학교 대학원 전자컴퓨터공학과

{jhchoi^o, luco, messias}@snoopy.korea.ac.kr, csjeong@charlie.korea.ac.kr

Service Oriented Workflow System on Grid

Ju-Ho Choi^o Yong-Won Kwon So-Hyun Ryu Chang-Sung Jeong
Dept. of Electronics and Computer, Korea University

요 약

그리드 어플리케이션을 분산, 병렬 환경에서 효과적으로 만들기 위해, 그리드 포털시스템인 Workflow based grid portal for problem Solving Environment(WISE)를 개발해왔다. WISE는 Globus Toolkit 2를 기반으로 그리드 워크플로우 기술을 사용하여 사용자가 좀 더 쉽게 그리드 어플리케이션을 프로그래밍 할 수 있는 환경을 제공한다. 그러나 현재 그리드 기술은 서비스 기반 구조(Open Grid Service Architecture)로 변화하고 있고 WISE도 기술적 흐름에 따라 그것을 충족시킬 수 있도록 변화해야 한다. 그래서 서비스 기반 구조에 대응할 수 있는 워크플로우 엔진을 새로 설계하여 WISE에 적용하고자 한다. 새로 설계한 시스템은 Service Oriented Workflow System on Grid(SOWS-G)이다. 이것은 계산량이 많은 문제에 대해 효과적으로 분산, 병렬 실행을 가능하게 하고, 다양한 워크플로우 패턴을 지원하며, 웹 서비스를 이용할 수 있도록 설계하였다. 이 논문에서는 SOWS-G를 서비스 기반 그리드 환경을 지원하는 Globus Toolkit 3를 이용하여 설계한 구조를 살펴보고 웹서비스를 이용할 수 있도록 추가된 기능들에 대해서 설명하고자 한다.

1. 서 론

그리드 컴퓨팅[1]은 이기종 분산 환경에서 다양한 자원을 효과적으로 이용할 수 있는 기술이다. 이 기술은 사용자가 분산된 자원에 대해 알 필요가 없기 때문에 많은 자원을 활용하여 프로그램을 쉽게 만들고 실행할 수 있는 환경을 제공한다. 따라서 그리드 컴퓨팅은 높은 계산 성능을 요구하는 e-Science 분야나 대용량 데이터 처리 등에 효과적으로 사용된다. 하지만 사용자에게 그리드 기술을 이용한 Problem Solving Environment(PSE)를 제공하기 위해서는 다양한 어플리케이션과 자원들을 제어할 수 있는 워크플로우 기술이 필수적이다.

과거 워크플로우 시스템들은 자신의 activity 만을 제어하고 사용하였으며 다른 워크플로우 시스템의 activity 와는 호환성이 없었다. 그러나 e-Business 환경에서는 워크플로우가 비즈니스 프로세스의 모델이 되고, 다른 activity 나 워크플로우간에 통신도 필요하게 되었다. 서비스 기반 환경(Service Oriented Environment)[2]은 이런 요건을 충족시켜준다. 이것은 분산 컴퓨팅 분야에서 새롭게 떠오르는 기술이며 이것의 주요 목적은 상호호환성이다. 이 기술은 서로 다른 툴을 이용하여 서로 다른 제작자가 만든 분산 어플리케이션들이 상호간 통신을 할 수 있도록 만들어준다. 이것을 이용하면 e-Business 같은 분야에서 어플리케이션들이 상호간에 서비스를 이용할 수 있게 된다.

최근에 그리드 컴퓨팅은 서비스 기반 환경으로 변화하고 있다. 웹서비스[3]는 서비스 기반 구조를 적용한 하나의 모델이며 그리드 컴퓨팅 역시 Open Grid Service Architecture(OGSA)[4]를 기반으로 웹서비스 인터페이스를 사용하는 형태로 변화하고 있다. 웹서비스 인터페이스는 Web Service Description Language(WSDL)[5]을 통해 서술되고 다른 사용자가 필요한 웹서비스를 찾는데 사용된다. Globus Toolkit 3(GT3)[6]는 OGSA와 WSDL 인터페이스를 지원하여 그리드 컴퓨팅을 서비스 기반 환경으로 구축하도록 해주는 미들웨어로서 Grid service, stateful service, lifetime management, notification 등의 기능도 함께 제공하고 있다.

따라서 그리드 컴퓨팅 환경에서 동작하는 워크플로우 시스템도 GT3를 이용하여 서비스 기반 환경으로 구축할 수 있다.

이 논문에서는 Service Oriented Workflow System on Grid(SOWS-G)의 구조를 제시하고 activity들이 다른 웹서비스 제공자를 이용하는 방법에 대해 설명하고자 한다. 또한, GT3를 사용하여 OGSA와 웹서비스 인터페이스를 갖는 그리드 환경에서 워크플로우 시스템 구현 방법에 대해서도 설명할 것이다. 2절에서는 기존의 워크플로우 시스템과 연관 기술들에 대해서 설명하고, 3절에서는 기존 시스템의 개선할 사항에 대해서 알아본다. 4절에서는 SOWS-G의 구조와 함께 구현방법에 대해서 설명하고자 한다. 마지막으로 5절에서 결과 및 향후 과제를 정리한다.

2. 관련연구

2.1 워크플로우 시스템

비즈니스 프로세스에 관한 워크플로우 기술은 WfMC[7]에서 정의하고 있으며 현재 수행되고 있는 다른 워크플로우 관련 프로젝트는 아래와 같다. Condor의 DAGman[8]은 DAG-based 그리드 워크플로우 기술이고 Pegasus는 이것을 이용하여 그리드 환경에서 abstract workflow를 정의할 수 있도록 만들었다. Pegasus[9]는 실행파일이나 사용가능한 자원을 이용하여 logical data를 physical data로 대응시킨다. Grid Services Flow Language(GSFL)[10]은 XML 기반으로 만들어진 워크플로우 서술 언어로써 OGSA 프레임워크를 기반으로 하는 그리드 서비스 환경에서 XML 스키마를 사용하여 워크플로우를 정의할 수 있다. 이것은 워크플로우 시스템은 아니지만 웹서비스 인터페이스를 정의할 수 있는 워크플로우 서술 언어라는 점이 특징이다. BioOpera[11]는 가상 연구실 환경을 구축하여 실험하기 위한 워크플로우 엔진이다. 이것은 클러스터 컴퓨터를 이용하여 대용량 어플리케이션을 처리하는데 쓰인다. 이것을 바탕으로 JOpera[12]가 개발되었는데, 이것은 visual composition language를 제공하여 task 간에 data flow 및

control flow를 정의할 수 있다. JOpera는 또한 Universal Description, Discovery, and Integration(UDDI)를 검색하여 웹 서비스를 이용할 수 있는 기능도 제공한다. 그러나 위의 워크플로우 시스템들은 OGSA와 웹서비스를 기반으로 하는 환경을 부분적으로만 충족시키고 있다.

2.2 WISE

Workflow based grid portal for problem Solving Environment(WISE)[13]는 clients, web application server, network of computing resources의 3-tier 구조로 되어 있다. middle tier 인 web application server는 그리드 서비스와 자원에 접근하기 위한 Grid-enabling 소프트웨어이다. 이것은 다층구조를 가지며 Model-View-Controller 디자인 패턴과 Commodity Grid(CoG)[14] 기술을 사용하여 flexible한 환경을 구축하고 있다. 웹서버의 어플리케이션 엔진은 클라이언트의 요청에 따라 적절한 작업을 수행하는 등의 전체적인 컨트롤을 담당하고 있다. 이런 작업은 그리드 서비스 인터페이스를 통해 이루어지며 결과를 사용자가 볼 수 있도록 클라이언트의 브라우저로 display 하는 기능을 가지고 있다.

3. 구현 목표

핵심적인 목표는 상호호환성(interoperability)이다. 이것을 위해 WISE의 워크플로우 엔진은 서비스 기반 환경을 충족시키는 새로운 구조를 필요로 한다. 또한, 다른 웹서비스 제공자들을 activity처럼 이용할 수 있도록 추가적인 component들을 설계해야 한다. 이 밖에도 sub-workflow 지원, activity의 동적 할당을 통한 성능향상 및 scalability 향상 등의 워크플로우 엔진 기능 개선도 함께 할 것이다. 이것을 통해 기존의 계산량이 많은 문제해결 뿐만 아니라 e-Business 분야에도 우리의 시스템을 적용시킬 수 있을 것이다.

4. SOWS-G 구조

4.1 overall architecture

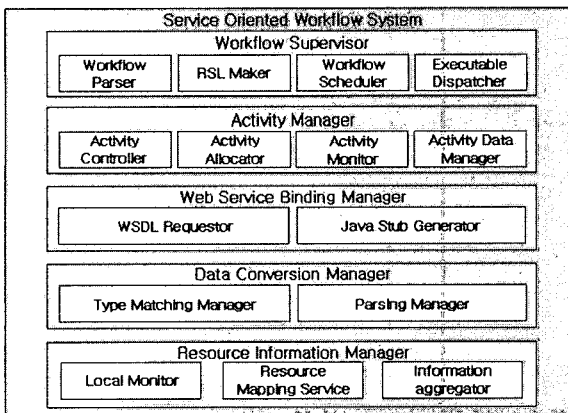


그림 1. SOWS-G 의 전체 구조

그림 1은 시스템의 전체적인 구조를 나타낸다. Service Oriented Workflow System on Grid(SOWS-G)는 WISE의 엔진에 해당한다. 이것은 Workflow Supervisor, Activity Manager, Web Service Binding Manager, Data Conversion

manager, Resource Information Manager 로 구성되어 있다. Workflow Supervisor는 워크플로우를 실행하기 위해서 서술된 언어를 파싱하고 실행하기 위한 스케줄링을 구성하며 작업할당에 사용될 명세를 만들어 준다. Activity Manager는 activity의 제어와 모니터링을 담당하고 있으며 이것을 Workflow Supervisor와 분리시켜 다른 노드에 독립적으로 할당할 수 있는 구조를 가지고 있다. 따라서 sub-workflow 개념을 사용하여 워크플로우를 만들 수 있으며, scalability 향상을 통한 성능 향상도 기대할 수 있다. Web Service Binding Manager는 웹 서비스 제공자의 주소를 통해서 WSDL 파일을 가져오고 그것을 통해 Java class를 생성하는 역할을 담당하고 있다. Data Conversion Manager는 웹서비스 제공자와 activity 사이에 data 교환 시 type check 및 matching을 해주는 부분이다. 마지막으로 Resource Information Manager는 local resource 의 상태를 모니터링 해서 모은 다음, 클라이언트의 요청이 있으면 적절한 자원을 알려주는 역할을 한다. 각 component 별 세부 사항은 아래에서 기술할 것이다.

4.2 Workflow Supervisor

Workflow Supervisor는 Workflow Parser, RSL Maker, Workflow Scheduler, Executable Dispatcher로 구성되어 있다. Workflow Parser는 WISE에서 사용하는 XML기반 워크플로우 서술 언어인 Grid Workflow Description Language(GWDL)[15]을 파싱한다. 그래서 activity와 node의 정보를 얻는다. 또한 GWDL은 pipeline, synchronization, parallelism 등을 통해서 AND-Loop, AND-Split, XOR-Split, XOR-Join 등의 다양한 워크플로우 패턴[16]을 지원하므로 이 부분도 여기서 처리된다. RSL Maker는 WDL로 부터 RSL 파일을 생성시킨다. RSL 파일은 Globus Toolkit에서 job을 제출하는데 사용된다. GT3에서는 XML 기반의 RSL이 필요하며 Globus Toolkit 2(GT2)에서는 기존의 RSL이 필요하므로 이 두 가지를 모두 지원한다. Workflow Scheduler는 Activity Manager로 activity의 정보를 보내 activity를 실행하도록 한다. 만약 GWDL에서 activity에 적절한 자원이 지정되어 있지 않으면, Resource Information Service의 Resource Mapping Service를 이용하여 자원을 할당받을 수 있다. Executable Dispatcher는 실행파일을 각 노드로 옮기는데 사용된다.

4.3 Activity Manager

Activity Manager는 Activity Controller, Activity Allocator, Activity Monitor, Activity Data Manager로 구성된다. Activity Controller가 나머지 세 개의 component를 제어한다. Activity Allocator는 job을 자원에 할당하고, Activity Data Manager는 data를 source site에서 destination site로 옮겨주는 역할을 한다. Activity Monitor는 activity의 state를 모니터링 한다. activity는 GT2, GT3, WS 세 가지 타입을 가진다. GT2 타입은 GT2를 기반으로 작동하며 GT2가 제공하는 기능인 GRAM을 통해 자원들을 이용하여 job을 할당하고 실행한다. 그리고 data의 이동은 마찬가지로 GT2가 제공하는 GridFTP를 사용하게 된다. GT3 타입은 GT3가 제공하는 OGSA 기반의 기능들을 사용하게 된다. XML기반의 RSL을 이용하는 WS-GRAM을 사용하여 job을 할당하고 실행하게 되며, RFT를 이용하여 옮길 데이터의 위치를 지정하면 RFT가 GridFTP를 이용하여 작업을 수행하게 된다. WS 타입은 웹서비스를 사용하는 activity를 나타낸다. SOAP을 이용하여 웹서비스 제공자와 연결을 하고 데이터를 주고, 받게 된다. 받은 데이터는 필요시, Data Conversion Manager로 보내져서 converting 이나 parsing 되

어 적절한 형태로 바뀌게 된다.

4.4 Web Service Binding Manager

Web Service Binding Manager 는 WSDL Requestor 와 Java Stub Generator로 구성되어 있다. WSDL Requestor는 웹서비스 제공자의 주소를 통해 WSDL을 가지고 온다. 그것을 이용하여 Java Stub Generator가 Java class로 바꾸어 준다. 사용자는 클라이언트 프로그램을 이미 만들어진 자바 클래스를 이용하여 쉽게 만들 수 있고 이것을 activity로 만들어 사용할 수 있다. 따라서 웹서비스를 activity로 만들어서 워크플로우를 구성할 수 있게 된다. 이 작업을 위해 Apache Axis 라이브러리를 사용한다.

4.5 Data Conversion Manager

Data Conversion Manager는 Type Matching Manager와 Parsing Manager로 구성되어 있다. 데이터 변환은 activity 사이, 또는 activity와 웹서비스 사이에 적절한 데이터 타입이나 내용을 보내기 위해서 필요하다. 기본적으로 제공하는 타입 변환기가 있으면 지정해서 사용할 수 있고 그렇지 않은 경우 사용자가 만든 타입 변환 프로그램이나 parser 등을 사용할 수 있다. 사용자는 GWDL 작성 시 이 부분을 추가하여 하나의 서비스로 사용할 수 있다. 사용자가 작성한 프로그램은 standard input/output을 따라야 하며, 처리한 결과를 Data Conversion Manager가 되돌려 준다. GT3 환경에서는 그리드 서비스를 정의할 수 있으므로 이미 정의된 서비스가 있다면 앞서 웹서비스를 사용하는 것처럼 WSDL 파일을 받아서 쉽게 사용자가 원하는 프로그램을 만들 수 있다. 따라서 사용자 의도대로 데이터를 처리할 수 있으며, 다양한 데이터 타입 변환을 통해 상호 호환성을 가질 수 있다.

4.6 Resource Information Manager

Resource Information Manager는 Local Monitor, Information Aggregator, Resource Mapping Service로 구성된다. Local Monitor는 CPU 사용량, 메모리 정보, OS Type 등의 정보를 모으는 역할을 한다. 그래서 이 정보들을 중앙에 있는 Information Aggregator로 보내어 모든 자원에 대한 정보를 실시간으로 파악한다. Resource Mapping Service는 GWDL 작성시, activity가 할당될 자원을 지원하지 않으면, 요청에 따라 현재 가용자원 중 가장 적절한 것을 자동으로 알려주는 역할을 한다. 이 기능을 통해 dynamic allocation을 구현할 수 있으며 계산량이 많은 일에는 성능향상도 기대할 수 있다. 이 기능은 GT3의 Web Service Information Service를 이용하여 구현된다. 로컬에서 자원의 상태를 파악하여 XML 형식으로 저장되고 Service data container가 그것을 수집하게 된다. 따라서 실시간으로 자원의 상태 파악이 가능해진다.

5. 결론 및 향후과제

이 논문에서는 서비스 기반 그리드 환경에서 작동하는 워크플로우 시스템(Service Oriented Workflow System on Grid)인 SOWS-G를 제시하고 각 구성요소별 기능을 설명하였다. 그리드 컴퓨팅 기술은 Open Grid Service Architecture를 바탕으로 Service-oriented 구조로 변화하고 있다. 여기에 분산 컴퓨팅의 새로운 모델인 웹서비스가 더해져서 발전해 나아가고 있다. Globus Toolkit 3는 이러한 환경을 구축해 주는 미들웨어이며, 우리는 이것을 바탕으로 SOWS-G를 설계하였고 현재 프로토

타입을 구현중이다. 따라서 지금까지 쓰였던 계산 집중형 문제뿐 아니라 비즈니스 모델에도 적용할 수 있을 만큼 보다 넓은 범위의 시스템으로 확장할 수 있었다. 또, 웹서비스를 이용하여 워크플로우를 구성할 수 있는 부분도 추가되었다. 아직은 GT3가 UDDI 기능을 지원하지 않기 때문에 불가능하지만, 향후에는 각각의 activity를 WSDL을 가지는 웹서비스 서비스의 하나로 할당하여 서비스 제공자의 역할까지 하도록 개발할 계획이다.

참고문헌

- [1] I. Foster, C. Kesselman, S. Tuecke. "The Anatomy of the Grid: Enabling Scalable Virtual Organizations.", International J. Supercomputer Applications, 15(3), 2001.
- [2] SOA and Web Services, <http://java.sun.com/developer/>.
- [3] W3C. Web Services, <http://www.w3.org/2002/ws/>.
- [4] I. Foster, C. Kesselman, J. Nick, S. Tuecke, Open Grid Service Infrastructure WG, Global Grid Forum. "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration." June 22, 2002.
- [5] Web Services Description Language (WSDL) <http://www.w3.org/TR/2004/WD-wsdl20-primer-20041221>.
- [6] Globus Toolkit 3, <http://www-unix.globus.org/toolkit/>.
- [7] Workflow Management Coalition, <http://www.wfmc.org>.
- [8] The Condor Project, <http://cs.wisc.edu/condor>.
- [9] Ewa Deelman, et.al, Pegasus: Mapping Scientific Workflows onto the Grid, Across Grid Conference, 2004.
- [10] Sriram Krishnan, Patrick Wagstrom, and Gregor von Laszewski. "GSFL: A Workflow Framework for Grid Services". Argonne National Laboratory, Preprint ANL/MCS-P980-0802, Aug 2002.
- [11] W. Bausch, C. Pautasso, R. Schaeppi, G. Alonso. "BioOpera: Cluster-aware Computing" Proceedings of the 4th IEEE International Conference on Cluster Computing, 2002.
- [12] Cesare Pautasso, Gustavo Alonso. "JOpera: a Toolkit for Efficient Visual Composition of Web Services", International Journal of Electronic Commerce (IJEC), 9(2):107-141, Winter 2004/2005.
- [13] Y.W. Kwon, S.H. Ryu, J.S. Park and C.S. Jeong, "A Workflow-Based Grid Portal for Problem Solving Environment", NPC 2004.
- [14] CoG Kits, <http://www.cogkit.org>.
- [15] Y.W. Kwon, S.H. Ryu, C.S. Jeong, H.W. Park, "XML-Based Workflow Description Language for Grid Application", ICCSA, pp 319-327, 2004.
- [16] Y.W. Kwon, S.H. Ryu, J.H. Choi and C.S. Jeong, "Workflow-Based Approach to Efficient Programming and Reliable Execution of Grid Applications", GCC 2004, LNCS 3251, pp.81-88, 2004.