

계산 e-사이언스 포털의 과학 워크플로우 서비스의 설계

김서영, 윤경아, 김윤희, 김종암
숙명여자대학교 컴퓨터과학과, 서울대학교 기계항공공학부
e-mail : {ssyyy77, yoonka, yulan}@sookmyung.ac.kr, chongam@snu.ac.kr

A Design of Scientific Workflow Service for an Computational e-Science Portal

Seoyoung Kim, Kyoung-a Yoon, Yoonhee Kim, ¹Chongam Kim
Dept. of Computer Science, Sookmyung Women's University,
¹Dept. of Mechanical and Aerospace Engineering, Seoul National University

요 약

계산과학공학 분야의 연구들은 빠르게 발전하는 컴퓨팅 기술에 비례해서 점점 더 컴퓨팅 기술에 의존하고 있으며 더욱 계산 집약적인 연구를 진행하고 있다. 최근 이러한 계산과학공학 분야에서는 별도의 컴퓨터 전문가를 요구하고 있고, 그들의 요구에 따라 실험에 특화된 컴퓨팅 인프라를 구성할 연구자는 턱없이 부족한 실정이다. 이에 따라 계산 과학 연구자들의 연구 환경 지원을 위하여 전 세계적으로 e-사이언스 환경에 대한 연구가 활발하게 진행되고 있으며, 여전히 연구자의 실험 환경과 컴퓨팅 자원 간의 별도의 설정과 설치 과정이 필요 없는 자유로운 연구 환경에 대한 보장이 요구되고 있다.

본 논문에서는 계산과학공학 분야의 연구자들의 실험 환경을 분산된 고성능 컴퓨팅 인프라와 함께 웹 브라우저를 통해 쉽게 접근 가능하고 또 사용하는 응용에 대한 수행이 가능하도록 통합된 실험 환경을 제시한다. 이 환경에서는 연구에서 주로 사용되는 워크플로우 중심의 수행을 제공하며, 이러한 수행을 위한 실험 설계를 가능하게 하는 워크플로우 설계 툴을 제공한다. 사용자는 각 계산에 필요한 수치 조건과 실행 방향을 drag-and-drop(드래그 앤 드롭) 방식으로 제공하는 각 모듈에 값만을 입력하여 원하는 형태의 실험을 진행 할 수 있으며, 이를 포털과 연결된 대규모 컴퓨팅 자원에 제출하여 그 결과를 빠른 시간 내에 제공받을 수 있다.

1. 서 론

계산과학공학 분야는 국가 과학 경쟁력을 높이고 산업 적용을 통하여 경제적인 경쟁력 강화에 중요한 역할을 하므로 전 세계적으로 매우 중요한 연구 분야로서 잘 알려져 있다. 하지만, 대규모의 고성능 계산 자원의 확충과 복잡한 설정이 요구되는 다수의 상용 소프트웨어에의 접근의 어려움으로 타 연구 분야에 비해 상대적으로 미흡한 발전을 보이고 있다. 이러한 문제점을 극복하고자 통합 실험 환경 연구가 최근 몇 년간 활발하게 진행되어왔다. 대표적 계산과학 연구 중 항공우주용 실험 환경인 e-AIRS [1]가 존재한다. e-AIRS는 교육용 사이버 연구 환경으로, 연구자 그룹 간의 효과적인 정보 및 협의 서비스를 제공해왔다. 그러나 주로 항공 우주 역학의 기초 교육 활용을 위해 설계되었기 때문에 사용자가 초보 연구자에 한정되어 있다. 상대적으로 복잡하고 다수의 값 비싼 소프트웨어를 사용하는 심화 연구자들은 오히려 직접 환경을 구축하여 실험을 해야 하는 상황이다.

본 논문에서는 항공우주 분야를 프로토타입으로 한 과학 워크플로우를 중심 통합 연구 환경을 제안한다. 워크플로우 중심 시스템은 전문가를 위한 복잡한 응용의 수행을

사용자가 정의한 흐름대로 자동화하여 효율적으로 실행해 주며 광범위하게 분산된 그리드 자원에서의 접근이 가능하게 하여 고성능 컴퓨팅 자원을 사용함으로써 단시간 내에 실험 결과를 제공받을 수 있다. 본 논문에서는 2장에서 관련된 연구, 3장에서는 포털의 서비스 구성에 대해 알아본다. 그리고 4장에서 워크플로우 설계 환경에 대해 설명하며 끝으로 결론 및 향후과제를 기술한다.

2. 관련 연구

과학 워크플로우 중심 문제풀이 환경에 대한 연구는 여러 계산 과학 분야에 걸쳐서 진행 되어왔다. 본 연구에서 대상으로 하는 항공우주 분야 외에도 생물학, 천문학, 물리학 등을 위한 워크플로우 설계를 지원하는 문제풀이 환경 연구로는 Taverna[2]과 Triana[3]가 있다.

Taverna[2]는 생물학을 중심으로 한 과학 워크플로우 관리 툴로, 이는 웹 서비스를 기반의 수행을 지원하며 함께 제공하는 워크벤치(Workbench)를 통해서 직접 웹 서비스에 연결하거나, 웹 서비스가 연결된 사이트를 통해 워크플로우 내에서의 사용을 가능하게 해준다. 또한 사용자는 가시적으로 표현된 패널을 이용해 각각의 프로세서, 링크들을 구성하여서 워크플로우를 설계할 수 있고 그 결과를 텍스트 파일 형태 또는 그래픽 툴에 연결되어 가시화 처

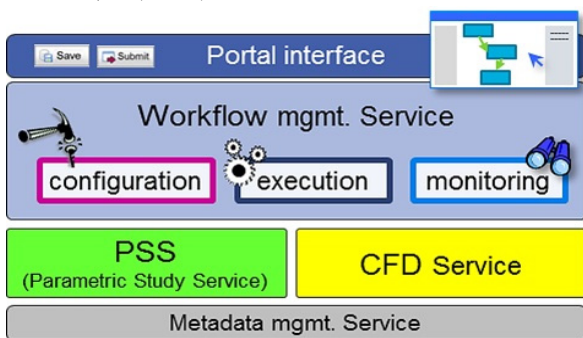
이 논문은 2010년도 정부재원(교육과학기술부)으로 한국연구재단의 지원을 받아 수행된 연구임 (No.2010-0027719)

리가 된 형태로 제공받을 수 있다.

Triana[3]는 천문학과 물리학을 대상으로 진행된 과학 워크플로우 구성을 지원하는 문제풀이 환경이다. 이는 그리드 서비스와의 연결을 제공하고 다양한 종류의 시스템들과 쉽게 통합 할 수 있는 조립-설치 가능한(pluggable) 컴포넌트들을 제공한다.

3. 기본 포탈 구조

본 논문에서는, 이전 연구[4]에서의 개선된 형태를 제공하며 다단계로 이루어진 확장된 형태의 응용의 적용에도 사용자가 쉽게 다룰 수 있는 형태의 인터페이스 위주로 통합된 환경을 제공한다. 이전 포탈 시스템에서 제공한 실험 서비스는 연구자에게 주로 고정된 실험 단계를 제공하거나, 간단한 옵션 선택 및 파라미터 값에 대한 확정 여부 기능이 제공되는 워크플로우 편집 서비스를 제공하였다. 하지만 더욱 확장된 응용을 적용하게 되는 경우, 전반적인 실험 흐름의 변경이 잦아 워크플로우 편집 서비스를 더욱 선호하게 되는 반면, 워크플로우 편집 서비스 내에서 제공되는 각 모듈에서의 속성 값 설정이 간단한 옵션 설정만이 가능하다는 한계점이 존재하였다. 이를 개선해 구체적인 실험 설정이 가능한 워크플로우 편집 서비스 형태를 구성하였다. 워크플로우 기반의 작업 수행 방식 및 데이터 관리는 기존과 동일하게 사용된다. 개선된 포탈 시스템의 구조는 그림1과 같다.



(그림 1) 포탈 시스템 서비스 구성

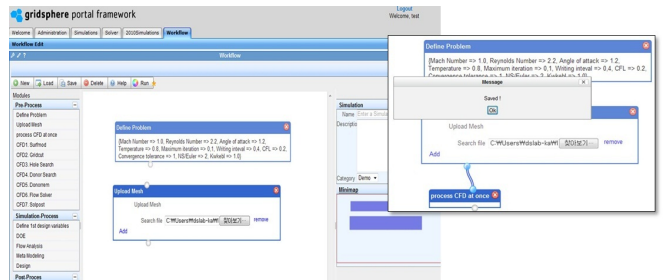
사용자는 워크플로우 편집기를 통해 실험을 구성하고, 작업의 순서와 속성 값들이 포함된 정보를 특정 시점에 포탈(인터페이스)로부터 전달 받게 된다. 전달 받은 데이터를 기반으로 '워크플로우 관리 서비스'는 컴퓨팅 자원 상에서 수행 가능한 형태로 워크플로우 구성(configuration)을 제공하며, 실행(execution) 서비스에서 작업 수행을 위한 스케줄링을 담당한다. 스케줄링 된 각 작업들은 워크플로우 모니터링(monitoring) 서비스에 의해 각 작업의 상태 정보와 워크플로우의 진행 사항을 사용자에게 제공한다.

이 밖에 워크플로우를 구성하는 작업들에 대한 다양한 범위의 파라미터 적용을 돕는 PSS와 각 작업들에게 요구되는 연산들의 제공과 생성된 결과에 대한 그래프 처리를 돕는 CFD 서비스는 이전 연구[4]와 동일하게 제공된다.

4. 워크플로우 설계 환경 구현

워크플로우 설계 환경의 구현에는 Wire-it[5]이 사용되었

다. 이는 Dataflow 응용 등에 사용되는 웹 인터페이스 기반의 오픈 소스 자바스크립트 라이브러리이다. 실험의 각 단계를 Wire-it에서 제공하는 각 모듈로 정의하였고, 모듈마다 실험 설정을 위한 파라미터 값 정의 및 여러 속성 값에 대한 구체적인 설계가 가능하도록 하였다. 각 모듈은 드래그 앤 드롭 방식으로 사용자 정의에 따라 배치하고 연결할 수 있으며, 작업 도중 작업 정보를 저장하기 위해서는 'Save' 버튼을 누르고, 최종 완성된 워크플로우를 계산 자원에 제출하기 위해서는 'Submit' 버튼을 누르면 된다. 실제 구현된 워크플로우 편집 서비스는 그림2와 같다.



(그림 2) 구현된 워크플로우 편집 서비스 모습

5. 결론 및 향후 과제

본 논문에서는 항공우주 응용에 특화된 워크플로우 중심의 문제풀이 환경을 제공하는 포탈 시스템을 제안하였다. 이러한 환경의 제공은 대규모 계산 자원의 확충 및 접근의 어려움으로 타 연구 분야에 비해 상대적으로 미흡한 발전을 보인 계산과학공학 연구자들에게 더욱 활성화 된 연구 환경을 제공할 수 있고, 활용도를 높여 고부가 가치의 상품과 기술을 쉽고 빠르게 만들 수 있으며 더불어 국가 과학 기술 경쟁력을 높일 것으로 예상된다.

향후 연구로 본 연구팀에서는 논리적 판단 및 비교가 가능한 멀티 스테이지 실험 환경으로 확장할 예정이다. 또한 효과적인 컴퓨팅 자원 제공을 위하여 그리드 환경 또는 클라우드 환경에 선택적으로 작업 제출이 가능한 고성능 컴퓨팅 환경 구성에 대한 연구를 진행하고 있으며 이러한 실험 환경과 함께 제공할 예정이다

참고문헌

- [1] 조정현, 허신영, 김은희, 김종암, 조금원, "e-Science 기반 사이버 교육을 위한 유체 해석 연구 시스템", KNOM Review, Vol. 12, No. 1, June. 2009, pp. 42-50
- [2] Oinn T, et al. "Taverna: A tool for the composition and enactment of bioinformatics workflows", Bioinformatics 2004; 20 (17):3045-3054
- [3] Majithia S, et al "Triana: A graphical web service composition and execution toolkit. IEEE International conference on Web Services (ICWS'04), SanDiego, U.S.A., 6-9 July, 2004
- [4] SY Kim, HJ Kang, KA Yoon, YH Kim, CY Hur, CA Kim, "Scientific workflow support of multi-stage experiments on an e-Scienceenvironment", ICONI 2010, pp.663~667, Dec. 16-20, 2010
- [5] Wire-it, <http://nevrlic.github.com/wireit/>