

국내 계산과학공학 커뮤니티를 위한 웹 기반 사이버-러닝 플랫폼 구축 및 서비스[☆]

Construction and Service of a Web-based Cyber-learning Platform for the Computational Science and Engineering Community in Korea

서 영 균^{1*} 조 금 원¹
Young-Kyoon Suh Kum Won Cho

요 약

최근 다양한 분야의 융합 연구에 대한 관심이 고조되고 있다. 이러한 시대 흐름에 발맞추어, 다수의 계산과학공학 분야의 연구자들의 연구 활동을 지원하고 학생들의 학습 의욕을 증진시키기 위한 IT기반 학제-간 융합 사업인 EDISON 프로젝트가 발주되었다. 우리는 2011년부터 이 EDISON 프로젝트를 성공적으로 수행해오고 있다. EDISON은 사이버-러닝 플랫폼으로서, 계산과학공학 연구자들이 거대 계산 및 I/O가 필요한 자신의 연구 문제를 해결하기 위해 개발한 고성능 컴퓨팅 시뮬레이션 소프트웨어를 웹에서 자유롭게 공유할 수 있게 하고, 사용자들의 아무런 제약 없이 실행할 수 있게 한다. 또한 EDISON은 국내 많은 대학의 강의 자료로 활용되고 있다. 본 논문은 이러한 EDISON 플랫폼의 구축 및 서비스 통계를 소개한다. 구체적으로, 우리는 기존 다른 고성능 계산과학 플랫폼과 EDISON 플랫폼과의 몇몇 차이를 설명하고, EDISON 플랫폼의 세 가지 기술적 계층 구조에 대해 논의한다. 이어 지난 4년 동안 EDISON 서비스에 대한 최신 통계를 제시한다. 끝으로, 본 논문의 결론을 맺고, 향후 계획에 대해 기술한다.

☞ 주제어 : 웹, 사이버-러닝, 플랫폼, 계산과학공학, 고성능 컴퓨팅, 시뮬레이션, 해석기

ABSTRACT

Recently, many attentions have been paid to conducting convergence research across diverse disciplines. Along with this convergence era, an IT-based multi-disciplinary convergence project, called EDISON (EDucation-research Integrated Simulation On the Net), has been launched to support the studies of researchers engaged in several computational science and engineering (CSE) fields and to boost learning motivations of CSE students. Since 2011, we have been successfully carrying out the EDISON project. EDISON as a cyber-learning platform enables CSE researchers to share their own high-performance computing (HPC) simulation softwares developed to solve their research problems accompanying large-scale computation and I/O and users to run the softwares with little constraints on the web. Also, the EDISON platform has been utilized as lecture material by many universities in Korea. This article introduces the construction and service statistics of this EDISON platform. Specifically, we explicate several distinctions between EDISON and existing other HPC service platforms and discuss a three-layered technical architecture of the EDISON platform. We then present the up-to-date service statistics of EDISON over the past four years. Finally, we conclude this article and describe future plans.

☞ keyword : Web, Cyber-Learning, Platform, Computational Science and Engineering, High-Performance Computing, Simulations, Solvers

1. 서 론

최근 학제-간 융합 연구에 대한 많은 사람들의 관심들이 고조되고 있다. 이러한 관심의 증가는 서로 다른 학문

간의 융합을 통해 새로운 연구 풍토를 조성하고 기존에 발견되지 않거나 알려지지 않았던 연구 기회를 얻기 위한 “연구 혁신”으로 이어지고 있다. 실제로, 최근 몇 년 동안 계산과학공학 분야는 컴퓨팅 능력, 네트워크, 저장 자원 등에서 진행된 고성능 컴퓨팅 기술의 획기적인 진보를 토대로 매우 활발한 연구를 진행해 오고 있다. 이러한 기술적 진보의 흐름을 타고 계산과학공학 연구자들은 거대 I/O가 요구되는 훨씬 계산-집약적인 문제를 해결하거나 사용자 친화적인 환경에서 자신들의 연구 결과를 통해 계산과학공학 분야의 학생들을 교육시킬 수 있는

¹ Supercomputing R&D Center, Korea Institute of Science and Technology Information (KISTI), Daejeon, 34141, Republic of Korea
* Corresponding author (yksuh@kisti.re.kr)

[Received 27 January 2016, Reviewed 25 February 2016(R2 April 2016., Accepted 20 June 2016)]

☆ 본 연구는 미래창조과학부 및 한국연구재단의 EDISON 사업 (NRF-2011-0020576) 지원으로 연구되었음.

성공적인 공유 방법에 큰 관심을 기울여오고 있다.

전형적으로, 계산과학공학 연구자들은 고성능 컴퓨팅 자원에 자신들이 개발한 시뮬레이션 프로그램* (이하 해석기)을 구동시키기 위해, 커맨드-라인 인터페이스를 통해 터미널에 접속한다. 하지만 다수의 계산과학공학 학생들이나 다른 연구자들은 터미널 상에서 일련의 명령어 입력을 통해 해석기를 실행하는 것에 친숙하지 않을 수 있다. 따라서 웹 포탈과 같은 사용자 친화적인 그래픽 유저 인터페이스를 이용한 분야-중립적 사이언스 게이트웨이[1]에 대한 강한 필요성이 제기되었다.

이에 따라, 2011년 우리 정부는 계산과학공학 및 IT 분야의 융합 사업인, EDISON (EDucation-research Integration through Simulation on the Net) 프로젝트를 발주하였다. EDISON은 다양한 계산과학공학문제를 해결하기 위해 개발된 다양한 해석기들을 공급하는 5개의 전문 센터—전산 열 유체, 나노 물리, 계산 화학, 전산 구조 동역학, 전산 최적 설계—와 이러한 해석기들의 유통 및 온라인 서비스를 담당하는 중앙 센터 간의 협력 프로젝트이다. 우리는 이 프로젝트를 수주하여 수년에 걸쳐 고성능 컴퓨팅 기반 온라인 계산과학공학 플랫폼으로서 EDISON을 설계, 개발, 서비스해 오고 있다. EDISON은 2012년 7월부터 2015년 6월까지, 5개 전문 분야 전국 43개 대학 661개 강좌 약 2만 8천여 명의 학생들이 수업에 활용해 오고 있는 국내 최대 규모의 이공계 사이버-러닝 플랫폼 중의 하나로 인정받고 있다 [2].

EDISON 플랫폼을 이용한 대표적인 교육 활용 사례로, 안동대학교의 수업을 들 수 있다. 이 수업에서 사용된 가상 풍동 실험은 기존에 사용되어 왔던 상용 SW가 아닌 EDISON에 탑재된 SW를 사용하여 진행되었다. 이를 통해 외부 라이선스 비용에 대한 지출을 줄일 수 있었으므로 비용 효율적 수업이 가능하게 되었다. 또한 EDISON을 활용해 본 학생들은 높은 흥미와 관심도를 가지고 수업에 임하게 되었고, 높은 수업 만족도를 얻었다 [2].

이러한 대중성을 바탕으로, EDISON 플랫폼은 2013년 IDC 주관의 고성능 컴퓨팅혁신 우수상 (아태평양 최초)을 수상한 바 있으며, 2014년 국내 기초연구 우수성과 50에 선정되는 성과를 거두었다. 이러한 EDISON 플랫폼의 기술적 우수성 및 국내외 인지도 상승에 힘입어, 지난 2015년 6월에 대만 고성능 컴퓨팅 연구소에 EDISON 전산 열 유체 포털을 수출하였다. 더욱이, 올해 후반에 베트

남의 계산과학연구소에도 EDISON을 제공할 예정이다.

본 논문은 EDISON 플랫폼의 구축 및 서비스 현황을 소개한다. 구체적으로, 우리는 EDISON 플랫폼의 핵심 구성 요소인 *애플리케이션 포털*, *미들웨어*, *인프라 스트럭처* (이하 인프라) 등 세 가지 기술적 계층 구조를 설명한다. 이러한 계층 구조의 조화를 통해 EDISON 플랫폼이 어떻게 1) (애플리케이션 포털 레이어의) 웹 포털을 통해 효율적으로 다양한 해석기들을 브라우징하거나 관리하는지, 2) 선택된 해석기를 통해 (미들웨어 레이어에서) 사용자들이 (인프라 레이어의) 고성능 컴퓨팅자원으로 온라인-시뮬레이션 작업을 수행 할 수 있는지 상세히 설명한다. 이어, 개발된 EDISON 플랫폼을 통해 서비스된 5개 분야 서비스 통계 자료를 제시한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서 기존 고성능 컴퓨팅서비스 플랫폼과 EDISON의 차이점을 설명하고, 3장에서 EDISON의 세 가지 계층 구조를 차례대로 상세히 기술한다. 4장에서 EDISON 서비스 현황을 제시한다. 5장에서 본 논문을 결론짓고 EDISON 플랫폼의 고도화를 위한 향후 연구 계획을 제시한다.

2. 관련 연구

표 1은 웹 기반의 다양한 고성능 서비스 플랫폼을 비교한다. 기존 고성능 서비스 플랫폼들[3, 4, 5, 6]은 EDISON과 같이 온라인 시뮬레이션을 지원하며 공공 서비스를 제공하는 선도 사이언스 게이트웨이로 인정받고 있다. 또한 기존 플랫폼들은 EDISON처럼 교육 및 연구 목적으로 사용되어 계산과학공학 커뮤니티 활성화에 기여하고 있는 사이버-러닝 플랫폼이다. 더욱이, EDISON을 포함한 이들 플랫폼의 시뮬레이션 서비스는 기본적으로 무료로 제공되고 있으나, nanoHUB [4]와 WebMO [5]등은 고급 기능이 포함된 상업용 버전도 제공한다.

그러나 몇몇 기존 플랫폼들 [3, 4, 5]은 몇 가지 한계점들을 드러냈다. 첫째, 이들 기존 플랫폼들은 지구 과학 [3], 나노 전자 [4], 계산 화학 [5] 등과 같이 특정 한 분야에 국한된 시뮬레이션 수행만을 지원하고 있다. 반면, HubZero [6] 및 EDISON은 특정 분야에 치우치지 않고, 타 계산과학공학 분야의 시뮬레이션 프로그램들을 유통할 수 있는 보다 진보된 개방형 범용 허브 플랫폼을 지향하고 있다. 또한 서로 다른 연구자들이 개발한 해석기 코드를 배포 및 공유함에 있어서, HubZero를 제외한 나머지 플랫폼들은 어떻게 공유 가능한지 기술하지 않고 있다.

* 소위 사이언스 앱 (Science App), 해석기 (solver), 혹은 실행 파일 (executable)이라고도 함.

(표 1) 다양한 웹 기반 고성능 컴퓨팅 서비스 플랫폼 비교
(Table 1) Comparison of Various HPC Service Platforms on the Web

	NEEShub [3]	nanoHUB [4]	WebMO [5]	HubZero [6]	EDISON
온라인 시뮬레이션	수행 가능	수행 가능	수행 가능	수행 가능	수행 가능
목적	교육 및 연구	교육 및 연구	교육 및 연구	교육 및 연구	교육 및 연구
유료 유무	무료	무료 (유료 버전 제공)	무료 (유료 버전 제공)	무료	무료
분야	지구과학	나노전자	화학	범용	범용
해석기 코드 배포 (공유) 방법	비공개	비공개	비공개	전용 툴킷 사용	관리자에 의한 승인 후 배포
국가	미국	미국	미국	미국	한국

HubZero는 해석기 코드 개발자들이 자신의 코드가 등록되고 배포되기 위해, 적지 않은 추가 업무를 부여한다. 즉 일명 랩처 (Rapture)라 불리는 전용 툴킷을 통해 XML 기반의 입력 스크립트를 해석할 수 있는 과서를 함께 제공함으로써 개발자들은 자신의 해석기 코드를 재포장해야 한다. 반면, EDISON은 개발된 해석기 코드가 인프라 자원에서 수행되지만 하면 관리자의 간단한 실행 테스트를 거쳐 바로 플랫폼에 등록되어 웹에서 서비스된다.

마지막으로, 기존 플랫폼들은 미국에서 개발되어 모든 서비스의 백-엔드 시스템이 미국에 위치해 있다. 국내 사용자들이 기존 서비스 플랫폼을 이용할 때 고속 네트워크 장비에도 불구하고 미국과 한국 사이의 지리적 원거리 때문에, 온라인 시뮬레이션 서비스의 질을 보장하기 어려울 수 있다. (전형적으로 큰 데이터 집합을 다루는 온라인 시뮬레이션을 수행함에 있어서 지연이 거의 없는 고속 네트워크를 지원하는 것이 매우 중요하다.) 그러나, EDISON은 모든 시스템에 국내에 위치하므로 이러한 네트워크 지연 시간에 대한 염려가 없다.

우리는 기존 플랫폼을 사용함에 있어서 위와 같은 여러 한계점들을 인식하였다. 따라서 우리만의 독자적인 범용 허브 플랫폼을 구축하기 위해 EDISON 플랫폼을 설계 및 개발하였다. 한걸음 더 나아가, 양질의 온라인 시뮬레이션 서비스를 국내의 사용자들에게 지원해 오고 있다.

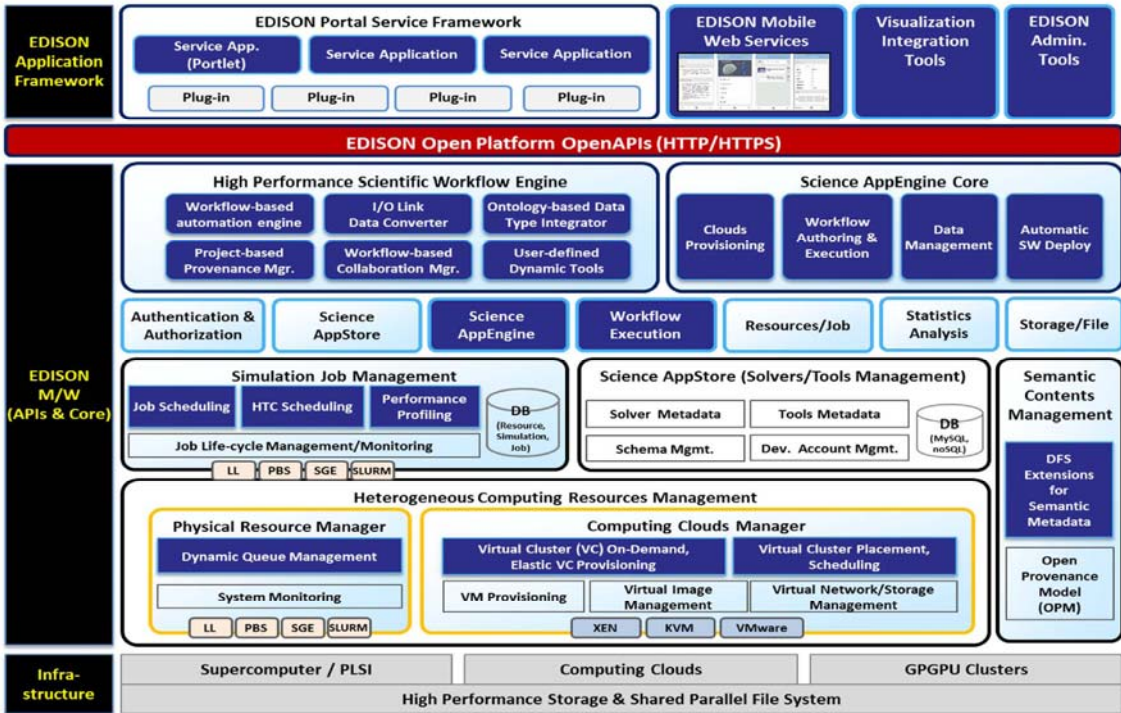
3. EDISON 구조

본 장은 EDISON 플랫폼의 각 구성 요소에 대해 논의한다. 그림 1은 EDISON의 전체 구조를 나타낸다. 그림 1에서 보는 바와 같이, EDISON 플랫폼은 크게 애플리케이션

프레임워크, 미들웨어, 인프라 등 세 가지 계층 구조로 되어 있다. (특히 그림 1에서 짙은 파란색으로 칠해진 블록 등은 현재 개발 중인 컴포넌트이다.) 애플리케이션 프레임워크는 포털 서비스로 구현되고, 포털 서비스는 포털릿의 조합으로 구현된다. 포털릿의 각 기능은 미들웨어에서 제공되는 API 등을 통해 구현된다. 미들웨어 레이어는 사이언스 워크플로우 엔진, 사이언스 앱 엔진, 시뮬레이션 작업 관리, 사이언스 앱스토어, 시맨틱 콘텐츠 관리, 이기종 계산 자원 관리 모듈 등으로 구성된다. 이들 미들웨어 컴포넌트들은 포털에서 수행되는 다양한 기능 (사용자 인증, 해석기 검색, 고성능 계산 작업 실행, 결과 데이터 다운로드 등)을 지원하기 위한 API를 제공한다. 인프라 레이어는 계산 작업 수행 및 입출력 데이터 저장을 위한 고성능 저장 및 공유 병렬 파일 시스템을 지원하는 이기종의 계산/저장 자원으로 이루어져 있다.

3.1 애플리케이션 포털 서비스 레이어

개방형 계산과학공학 시뮬레이션 서비스가 이용 가능하도록 하려면, 고성능 컴퓨팅서비스 플랫폼은 다음의 세 가지 특성을 지녀야 한다. 첫째, 지리적으로 떨어진 연구자들이 서로 인-하우스 해석기 도구들 및 데이터 집합을 공유하고 제약 조건 (라이선스 구매 등) 없이 어떤 사용자든지 게시된 해석기 도구 등을 사용할 수 있어야 한다. 둘째, 그 연구자들이 쉽게 도구들을 공유하고 실행해 볼 수 있도록 사용하기 쉬운 직관적인 그래픽 유저 인터페이스를 제공할 수 있어야 한다. 셋째, 범용 플랫폼으로서 가치를 지니기 위해 더 많은 다양한 계산과학공학 분야를 포용할 수 있도록 쉬운 이식성 (portability)을 지닐 수



(그림 1) EDISON 플랫폼 전체 구조
(Figure 1) The overall architecture of the EDISON platform

있어야 한다. 그러므로 위의 세 가지 조건들을 만족시킬 수 있는 가장 적합한 해결책으로 우리는 “웹-기반 문제-해결” 환경을 고려하게 되었다.

Liferay[7] 포털 프레임워크는 위의 세 가지 특성을 가장 잘 반영하는 웹 기술 프레임워크이다. Liferay는 EDISON 사용자에게 보이는 다양한 포틀릿을 지원하는데, 이러한 포틀릿은 EDISON 플랫폼의 개방형 API를 통해 각종 미들웨어 서비스들을 접근할 수 있다. 이 API는 웹 표준 RESTful (REpresentational State Transfer Full) 인터페이스 [8]를 따른다.

그림 2는 EDISON의 5개 분야 중 전산 열 유체 사이트 (<http://cfd.edison.re.kr>)를 보여준다. 그림 2와 같이, 사용자는 ‘Simulation’ 메뉴를 이용하여 다양한 해석기 (사이언스 앱)들을 브라우저 할 수 있다. 이후, 원하는 해석기를 선택하고, 필요한 파라미터 값들을 입력하여 실제 온라인 시뮬레이션을 수행할 수 있다.

그림 3은 EDISON 모바일 서비스를 나타낸다. 모바일 서비스는 위에서 언급된 동일한 RESTful API들을 메뉴

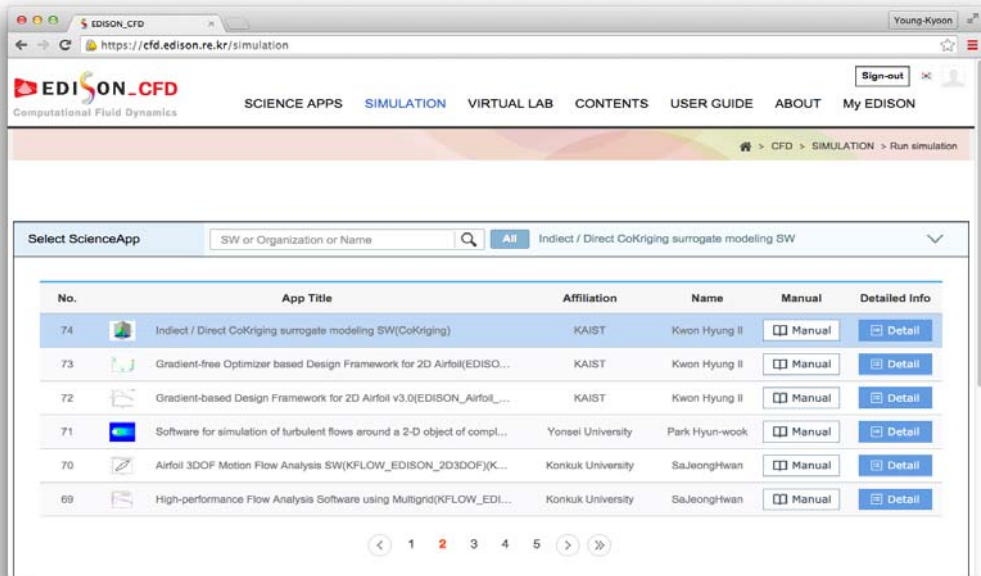
및 기능 별로 호출함으로써 구현된다. 이를 통해, EDISON 포털 사이트에서 제공되는 모든 기능을 사용자들이 모바일 웹에서도 동일하게 사용할 수 있도록 한다.

3.2 미들웨어 레이어

EDISON 미들웨어 레이어는 사용자 인증 및 권한, 해석기 메타데이터, 시뮬레이션 데이터셋, 시뮬레이션 이력 관리, 사이언스 워크플로우 실행, 이기종 (가상 및 물리적) 계산 자원 관리 등의 다양한 플랫폼 핵심 기능을 수행한다. 미들웨어 레이어는 크게 1) 메타데이터 관리, 2) 가상화 자원 및 작업 관리, 3) 사이언스 워크플로우 관리 프레임워크 등 세 가지 구성요소로 이루어져 있다.

3.2.1 메타데이터 관리 프레임워크

일반적으로 계산과학공학 분야 시뮬레이션 환경은 입력 제어를 위한 전처리기, 계산과학 해석기, 시뮬레이션 결과의 시각화를 위한 후처리기 등으로 나뉠 수 있다.



(그림 2) EDISON 열유체 포털
(Figure 2) The EDISON CFD portal site

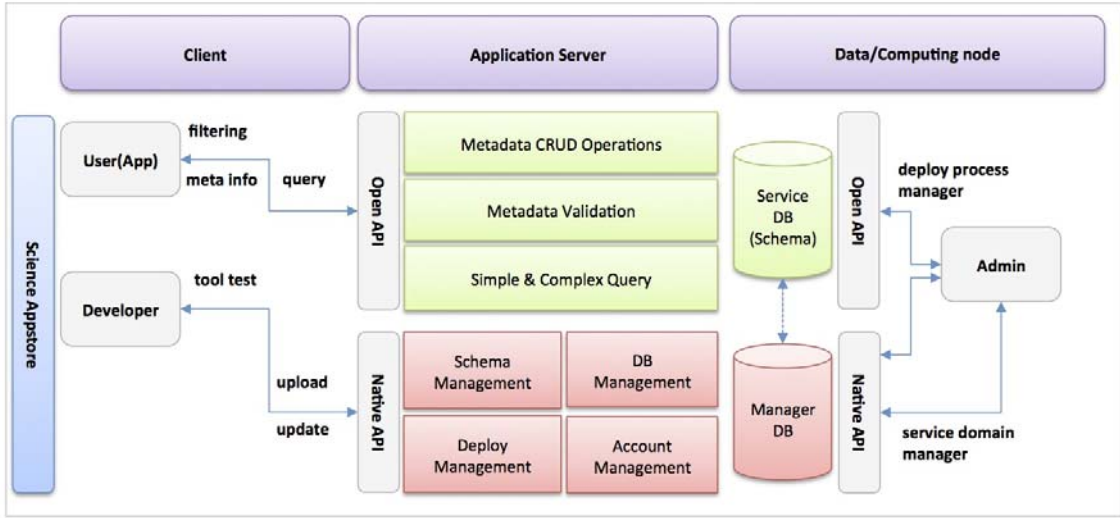


(그림 3) EDISON 모바일 서비스
(Figure 3) EDISON mobile service

이들 구성 요소를 활용하는 사이버-인프라스트럭처를 구축할 때 한 가지 이슈는 이러한 구성 요소들에 접근할 “표준화된 인터페이스”가 없다는 것이다. 또 다른 이슈는 이들 구성 요소들에 대한 속성들이 유연하게 변경 가능할 수 있어야 한다는 것이었다.

이러한 이슈들을 해결하기 위해, 우리는 EDISON 메타데이터 프레임워크인 사이언스 앱스토어를 설계하고 개발하였고, 그 구조는 그림 4와 같다. 사이언스 앱스토어는 메타데이터 저장소 및 메타데이터 등록 및 관리를 위한 API 세트를 지원한다. 앱스토어의 메타데이터 저장소는 해석기와 연관된 메타데이터의 유연한 확장성을 고려해 NoSQL (Not-only SQL) 기반의 데이터 관리 시스템인 MongoDB [9]를 통해 구현되었다. 앱스토어에 저장된 메타데이터는 Node.js [10]를 사용하여 세워진 RESTful 인터페이스로 구성된 *SpyGlass*라 불리는 관리자에 의해 접근되고 관리된다.

그림 4에서 보는 바와 같이, 사이언스 앱스토어의 개방 API는 해석기 메타데이터에 대한 CRUD (create/read/update/delete) 연산 및 유효성 검증을 수행한다. 사용자는 *SpyGlass*를 통해 (사이언스 앱스토어의 접근 API를 호출함으로써) 자신이 지정한 필터링 조건을 만족시키는 해당 해석기를 검색해 낼 수 있고 선택된 해석기에 대한 시



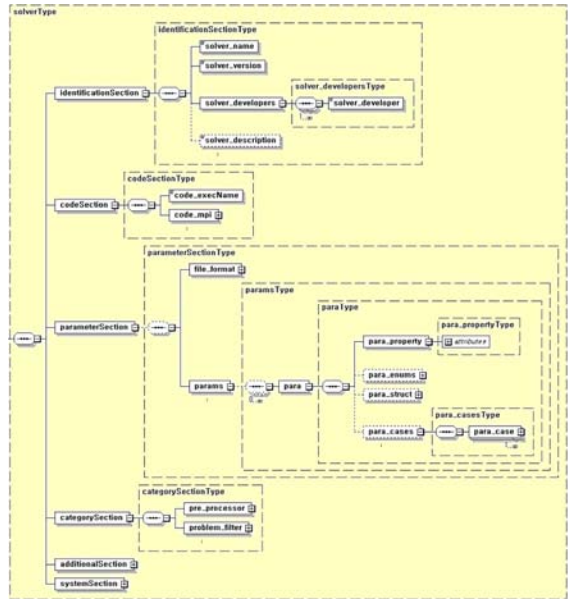
(그림 4) 사이언스 앱스토어 구조
(Figure 4) The architecture of Science AppStore

물레이션 실행 정보를 알아낼 수 있다.

이외에도, 사이언스 앱스토어는 개발자가 자신이 개발한 해석기를 등록하고 배치할 수 있는 원천 API도 제공한다. 즉 개발자는 이 API를 이용하여 자신의 해석기에 대한 기본 정보 (이름, 특징, 설명 등) 및 시물레이션 실행 정보 (입력 파라미터, 출력 결과 파일 등)와 함께 테스트된 해석기 코드 위치 등을 입력하면, 사이언스 앱스토어가 이를 저장하고 이렇게 등록된 해석기 정보는 EDISON 관리자에게 전달된다. 관리자는 해당 정보를 바탕으로 실제 데이터 및 계산 노드에 등록 요청된 해석기를 실행하고 이상이 없으면, 최종적으로 EDISON 플랫폼에 서비스 된다.

우리는 5개의 전문 분야에서 사용되는 다양한 해석기들을 분석하여, 그림 5에서 보는 바와 같이 전체 79개의 엘리먼트, 53개의 속성으로 이루어진 글로벌 스키마를 설계했다. 이러한 글로벌 스키마는 크게 6 개의 섹션으로 구분된다. 먼저, 1) 식별 섹션은 해석기의 제목, 버전, 개발자, 소속 등과 EDISON 사용자들을 위한 설명, 특징, 스크린샷 등 다양한 식별 정보를 포괄한다. 2) 코드 섹션은 자동화된 작업 명령어를 생성하기 위한 정보를 담는다. 이러한 정보는 해석기 파일 이름, 경로, 결과 파일 경로 등과 링커를 위한 프로그래밍 언어, 컴파일러, 정적 라이브러리 등을 포함한다. 3) 파라미터 섹션은 해석기의 입력으로 제공되는 제어 파라미터의 이름, 데이터 타입, 설명, 기본값 등 각종 파라미터에 대한 정보를 포괄한다.

4) 카테고리 섹션은 시물레이션 실행 프로세스에 참여하는 구성요소의 구분 (전처리기, 해석기, 후처리기 등)에 대한 정보 및 사용자들이 즐겨 찾는 해석기 정보를 포괄한다. 5) 추가 섹션은 해석기와 관련된 콘텐츠 (기술 보고



(그림 5) 시물레이션 소프트웨어를 위한 글로벌 스키마
(Figure 5) The Global Schema for Simulation SWs

서, 연구 논문, 강의 자료 등)를 포괄한다. 6) 시스템 색서는 EDISON 플랫폼에서 수행된 해석기에 대한 관리자 정보 (버전 제어, 사용 빈도, 서비스 상태 등)를 포괄한다. 해석기에 대한 메타데이터는 이러한 글로벌 스키마를 준수하는 문서형 형태로 표현되어 JSON 파일로 MongoDB에 저장되어 관리되고 검색된다.

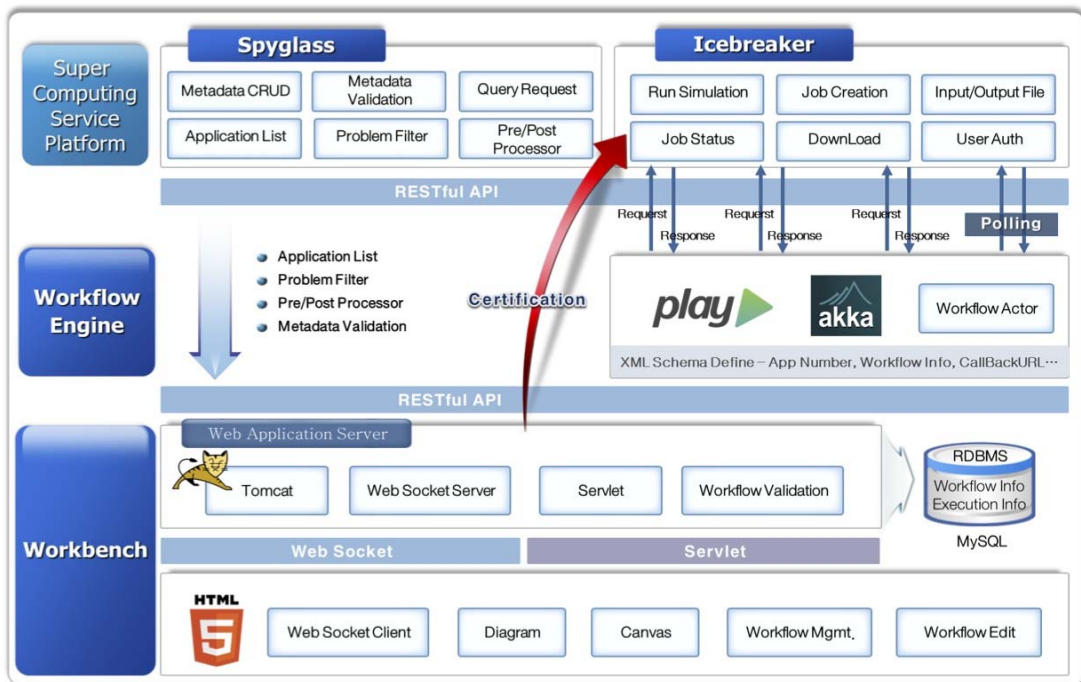
3.2.2 가상화 작업 및 작업 관리 프레임워크

EDISON 플랫폼의 가상화 계산 자원 및 자원 관리 프레임워크는 이용 가능한 계산 자원의 효율적인 사용과 시뮬레이션 작업 및 자원을 쉽게 모니터링할 수 있도록 하는 웹-표준 인터페이스 등을 고려하여 설계되었다. 이 프레임워크의 핵심인 소위 Icebreaker라는 작업 및 자원 관리 모듈은, 1) 사용자 인증, 자원 가상화, 시뮬레이션 작업 및 자원 관리 등을 위한 추상화 레이어, 2) 로컬 DB 및 LDAP [11]을 이용한 사용자 관리, Xen을 이용한 (물리적/가상) 서버 프로비저닝 및 OpenPBS 및 네트워크 파일 시스템 등을 이용한 시뮬레이션 작업 관리 및 실행을 위한 실제 서비스 레이어, 3) 가상화 자원 및 시뮬레이션

작업에 접근하거나 관리하기 위한 시스템 관리자를 위한 웹-서비스 레이어 등으로 구성된다. 특히, 사용자 정보, 물리적 서버 (호스트), 가상머신, 가상 네트워크, 스토리지, 시뮬레이션 작업 등은 Hibernate/HSQL [12]과 결합된 플레인 자바 오브젝트 (POJOs)로 추상화 되어 관리 된다. 이들 플레인 자바 오브젝트에 접근하기 위한 CRUD 연산은 RESTful 인터페이스를 이용하여 Spring [13] 프레임워크 기반의 웹-서비스 레이어를 통해 구현되었다.

3.2.3 사이언스 워크플로우 관리 프레임워크

EDISON 플랫폼은 또한 계산과학공학 분야의 중요한 요청 사항인 사이언스 워크플로우를 지원한다. 그림 6은 EDISON 워크플로우 프레임워크인, SimFlow [14]의 전체 구조를 나타낸다. 사용자들은 SimFlow를 통해 서로 다른 해석기들의 사이언스 워크플로우를 구성하고 실행할 수 있다. SimFlow는 크게 두 가지 구성 요소인 워크플로우를 저작하고 저장하기 위한 자바 기반의 1) 워크벤치와 저장된 워크플로우를 이 워크벤치 내에서 실행하는 2) 실행 엔진으로 이루어져 있다. 워크벤치는 SpyGlass를 통



(그림 6) EDISON 워크플로우 실행 프레임워크 구조 [14]

(Figure 6) The architecture of the EDISON workflow framework (SimFlow) [14]

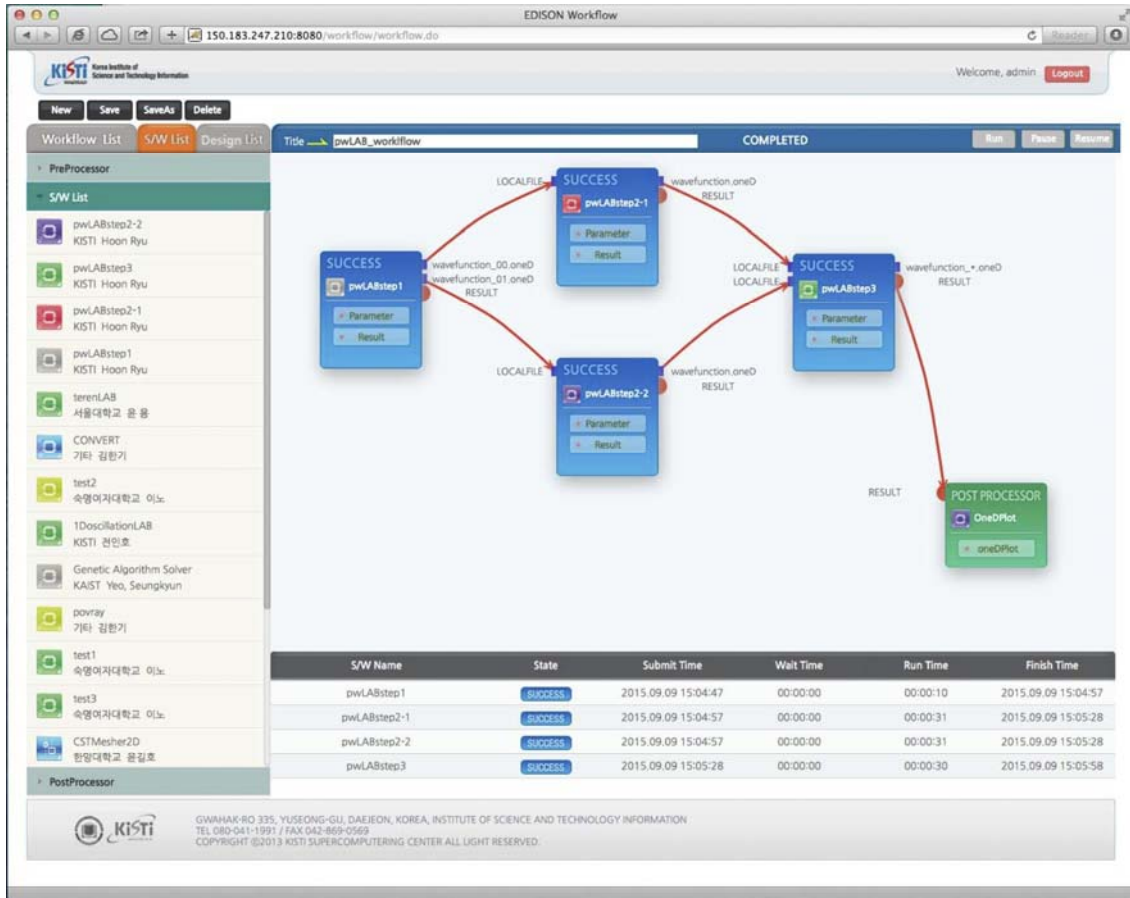
해 해석기 및 관련 메타데이터를 얻어올 수 있으며, 사용자들의 워크플로우 구성을 위한 편집 기능을 제공한다. 워크벤치는 Spring, jQuery, jsPlumb 등을 이용하여 구현되었다. 실행 엔진은 저장된 워크플로우의 실행을 위해 Icebreaker와 연동하여 동작한다. 워크플로우의 동시적 실행을 위해 Play 프레임워크 및 Akka 기술이 이용되었다.

그림 7은 웹 기반 사이언스 워크벤치를 보여준다. 왼쪽 메뉴를 이용해, 사용자는 이용 가능한 해석기 리스트를 볼 수 있다. 원하는 해석기를 선택하여 드래그-앤-드랍 방식을 이용해 배치할 수 있다. 이때, 입력 및 출력 파라미터의 타입이 서로 동등한 해석기들만 연결 가능하다. 이어, 사용자는 작성된 사이언스 워크플로우를 저장하고 실제로 실행해 볼 수 있다. 일단 워크플로우가 실행되면, 그림 7의 하단에 보이는 바와 같이 각 해석기에 의한 시

물레이션 상태 정보가 표시된다. 해당 워크플로우의 실행이 성공적으로 종료되면, 후처리기 (예. oneDPlot) 등을 이용해 최종 결과를 다운로드하여 가시화 할 수 있다.

3.3 인프라 레이어

EDISON 인프라 레이어는 사용자가 수행한 시뮬레이션 작업에 의한 대규모 계산 및 입출력을 지원하기 위해 설계되었다. 계산 자원은 크게 슈퍼컴퓨터 및 국가슈퍼컴퓨팅 공동 활용 체계, 1,168개의 물리적 코어로 구성된 계산 클라우드 (클러스터) 및 GPGPU로 구성되어 있다. 다수의 사용자에 의한 대량의 입출력 트래픽을 처리하기 위해, 고가의 상용 스토리지 장비를 사용하지 않는 대신 비용-효과적으로 처리할 수 있는 두-계층에 의한 저장 자



(그림 7) 작성된 사이언스 워크플로우 실행
(Figure 7) The EDISON workflow framework (SimFlow)

원 아키텍처를 구축하였다. 즉 총 12TB의 용량을 보유한 두 대의 고성능 SSD 노드에 확장형 네트워크 파일 시스템으로 잘 알려진 GlusterFS [15]을 설치하여, 이를 “프런트-엔드”로 비치하였다. 또한 SSD 노드의 가용 저장 공간이 부족할 경우를 대비하여, “백-엔드”로 총 48 TB에 이르는 하드디스크 드라이브 기반의 네트워크 파일 시스템을 구축하였다. 이러한 두-티어 구조를 통해, 5개 전문 분야에서 사용 중인 다양한 시뮬레이션 프로그램 도구들의 시뮬레이션 작업으로 파생된 무거운 I/O 트래픽을 효과적으로 처리해오고 있다.

4. EDISON 서비스 현황

우리는 인프라 구축 경험 및 EDISON 플랫폼을 바탕으로, 지난 4년 동안 국내 계산과학공학 분야 이공계 학생들에게 웹 기반 시뮬레이션 사이버-러닝 서비스를 제공해 왔다. 표 2는 이에 대한 구체적인 통계 자료 [16]이다.

EDISON을 활용하고 있는 대학의 수가 가장 많은 분야는 전산 열 유체였다. 이는 EDISON 프로젝트가 시작된 이래로 가장 오랫동안 참여해 왔으므로 그동안 EDISON 활용에 대한 홍보가 많은 대학들로 이어진 결과로 생각된다. 개설 강좌 수 및 활용 학생 수는 계산화학이 가장 많았다. 이는 과제 참여 교수님의 분포가 다양했기 때문에 개설된 강좌에 참여한 학생 수도 자연스럽게 늘어난 것으로 생각된다. 제출 작업 수는 나노 물리 분야가 가장 많았는데, 이는 파라미터 스윕에 의한 작업 수행이 많았기

때문이었다. 가장 많은 작업 계산 시간이 소비된 분야는 전산 열 유체였다. 이는 전산 열 유체 분야에 탑재된 많은 해석기들이 계산-집약하기 때문으로 풀이된다.

이러한 결과는 다음의 세 가지 시사점을 보여 준다. 첫째, EDISON은 다수의 계산과학분야에서 수행되는 대규모 시뮬레이션 작업을 수행하기에 적합한 프레임워크임을 입증한다. EDISON은 누적 합계 3만 여명에 이르는 사용자들이 제출한 84만 여건의 작업에 대해 약 73만 시간에 이르는 거대한 계산을 안정적으로 지원하였다. 둘째, 분야별로 평균 46 여개의 대학, 132 여개의 강좌, 약 5 천여 명의 사용자를 갖는 대중성을 확보하였다. 이는 다중 분야에서 적용 가능한 EDISON의 기술적 우수성을 반증하며, 범용 계산과학 시뮬레이션 프레임워크로의 가능성을 보여 준다. 셋째, 각 분야에 특화된 서비스 정책을 고려할 필요가 있다. 제출 작업 수 및 계산 시간이 분야별로 상이한 분포를 보인다. 따라서 분야 별 우선 작업 스케줄링 또는 계산 자원 할당에 고려가 필요할 것으로 보인다.

5. 결론 및 향후 계획

본 논문은 계산과학공학 연구자 및 학생들을 위한 온라인 시뮬레이션 서비스를 제공하는 EDISON 플랫폼 및 서비스 현황을 소개하였다. 본 논문은 EDISON의 세 가지 계층 구조 및 각 계층의 구성 요소에 대해 상세히 설명하였다. 융합이라는 흐름에 따라, 원자력, 생명정보학, 전산 재료, 전산 수학, 신약 등 더 많은 분야에 대한 온라

(표 2) EDISON 서비스 누적 통계 [16]

(Table 2) Cumulative Statistics of the EDISON Service [16]

적용 분야	활용 대학 수	개설 강좌 수	활용 학생 수	제출 작업 수	컴퓨팅 시간 (시)
전산열유체	118	242	7,781	66,071	450,068
나노물리	45	135	7,026	624,066	85,235
계산화학	57	270	12,579	129,197	187,871
구조동역학	7	9	389	16,432	9,887
전산설계	4	5	277	8,148	305
총계	231	661	28,052	843,914	733,366
평균	46	132	5,610	26,871	146,673

인 시뮬레이션 서비스를 지원함으로써, EDISON이 다재 다능함을 확장하고 더 높은 대중성을 얻는 차세대 연구 플랫폼으로 발전할 수 있을 것임을 확신한다.

물론 앞으로 EDISON은 다음과 같은 추후 연구를 통해 그 기능을 강화할 필요가 있다. 먼저 현재는 사용자들이 선택된 시뮬레이션 프로그램에 대해 동일한 입력을 제공할 경우, 중복해서 시뮬레이션을 실행하도록 되어 있다. 이는 계산 및 저장 자원에 대한 효율성이 저하되는 문제를 노출한다. 따라서 동일 또는 유사 입력에 대한 시뮬레이션 작업의 경우 기 실행된 결과가 존재한다면, 실제 실행 없이 쌓여있는 데이터를 기반으로 한 데이터-주도형 시뮬레이션을 수행할 수 있는 서비스 기술에 대한 적용이 필요하다. 또한 사용자에 의해 제출된 계산 작업에 대해 좀 더 현명한 스케줄링 기법이 필요하다. 구체적으로, EDISON은 현재 순차적 기반의 라운드-로빈 방식을 적용하고 있다. 하지만, 특정 작업이 빨리 종료될 수 있거나 늦게 종료될 수 있는 것으로 EDISON에서 예측될 경우, 이들의 작업 순서를 동적으로 재배치함으로써 작업 처리량을 상승시킬 수 있는 새로운 스케줄링 기법을 고안하고 적용시킬 필요가 있을 것으로 보인다.

참 고 문 헌 (Reference)

- [1] Y.-K. Suh, et. al., "EDISON: A Web-Based HPC Simulation Execution Framework for Large-Scale Scientific Computing Software," in Proceedings of the ACM/IEEE CCGrid'16, May 2016, pp. 608-612.
- [2] Inho Jeon et. al., "Development and Application of Smart-Learning Technology in Science and Engineering Fields," KIPS Review, Vol. 23, no. 1, 2016, pp. 16-22. http://society.kisti.re.kr/sv/SV_svpbs03V.do?method=view&art_no=1&id=0&iSet_id=82&totalCount=1&menuid=1&subid=11
- [3] T. J. Hacker, et. al., "The NEEShub Cyberinfrastructure for Earthquake Engineering," Comp. in Sci. & Eng., Vol. 13, no. 4, 2011, pp. 67-78.
- [4] G. Klimeck, et. al., "nanoHUB.org: Advancing Education and Research in Nanotechnology," Comp. in Sci. & Eng., Vol. 10, no. 5, 2008, pp. 17-23.
- [5] J.R. Schmidt and W.F. Polik, WebMO Enterprise, version 13.0; WebMO LLC: Holland, MI, USA, 2013; <http://www.webmo.net> (accessed March, 2013).
- [6] M. McLennan and R. Kennell, "HUBzero: a Platform for Dissemination and Collaboration in Computational Science and Engineering," Comp. in Sci. & Eng., Vol. 12, no. 2, 2010, pp. 48-53.
- [7] Liferay. Liferay Portal Framework. <http://www.liferay.com>, (accessed September, 2015).
- [8] C. Pautasso et. al., "Restful Web Services vs. "Big" Web Services: Making the Right Architectural Decision," in WWW '08, pp. 805-814. ACM, 2008
- [9] MongoDB, Inc. MongoDB. <https://www.mongodb.org/> (accessed January 2016).
- [10] Node.js Foundation. Node.js. <https://nodejs.org> (accessed January 2016).
- [11] T. Howes and M. Smith, LDAP: Programming Directory-enabled Applications with Lightweight Directory Access Protocol, Macmillan Publishing, 1997.
- [12] JBoss.org, Database Configuration for Hibernate, <https://docs.jboss.org/executor/1.12.13-GA/developer/en-US/html/ch-db-configuration-hibernate.html> (viewed on April 2016).
- [13] Spring Framework, <http://www.springframework.org> (viewed on April 2016).
- [14] H. Kim et al., "A Design and Implementation of a Lightweight Scientific Workflow System for Higher Education and Advanced Research of Computational Science and Engineering," Information, Vol. 17, no. 11, 2014, pp. 5927-5932.
- [15] Gluster, GlusterFS. <https://www.gluster.org> (accessed January 2016).
- [16] K. W. Cho et. al., "The 4th Annual Report of the EDISON Project," June 2015.

● 저 자 소개 ●



서 영 균 (Young-Kyoon Suh)

2003년 경북대학교 컴퓨터학과 (현, 컴퓨터학부) 졸업 (학사)

2005년 KAIST 전자전산학과 전산학 전공 졸업(석사)

2015년 Dept. of Computer Science, Univ. of Arizona 졸업(박사)

2005년~현재 한국과학기술정보연구원 (KISTI) 슈퍼컴퓨팅연구센터 선임연구원

관심분야 : 데이터베이스 시스템/설계, 컴퓨팅의 과학, 빅데이터, HPC, 소프트웨어 테스트

E-mail : yksuh@kisti.re.kr



조 금 원 (Kum Won Cho)

1993년 인하대학교 항공우주공학과 졸업(학사)

1995년 KAIST 항공우주공학과 졸업(석사)

2000년 KAIST 항공우주공학과 졸업(박사)

2000년~현재 한국과학기술정보연구원 (KISTI) 슈퍼컴퓨팅연구센터 책임연구원 (센터장)

관심분야 : 계산과학, 항공우주, 유체해석, 이공계 교육·연구 융합

E-mail : ckw@kisti.re.kr