

React 기반 시뮬레이션 이력 데이터 서비스 프레임워크 설계 및 구현

Design and Implementation of a React-based Simulation-Provenance Data Service Framework

요 약

본 논문은 새로운 웹 기반 고성능 컴퓨팅(HPC) 시뮬레이션 이력 데이터 서비스 프레임워크 SantaFe를 소개한다. 이 프레임워크는 시뮬레이션 이력 데이터에 관한 두 가지 핵심 서비스를 제공한다. 첫째, 사용자가 웹 상에서 입력한 인자에 대한 시뮬레이션 실행 종료 시간을 사전 학습된 기계 학습 모델을 통해 수초 내에 추정하여 보여준다. 둘째, 특정 시뮬레이션 프로그램에 입력되었던 파라미터 집합 정보를 빈도 순으로 나열하여 보여준다. 프레임워크는 React 라이브러리 및 Node.js를 활용하여 구현되었으며, R로 개발한 시간 추정 모델을 웹에서 동적으로 호출할 수 있게 하였다. 이 프레임워크를 통해 계산 과학 HPC 시뮬레이션을 온라인에서 실행하는 사용자의 편의성이 크게 개선될 수 있을 것으로 기대된다.

1. 서 론

최근 물리, 화학, 생물 등 다양한 연구분야에서 컴퓨터를 이용한 다양한 수치적 모델링 및 계산 과학 연구가 진행되고 있다. 계산 과학의 경우 직접 실험이 불가능한 현상을 컴퓨터를 통해 모사하고 가시 영역 이하의 물질 거동을 체계적으로 연구할 수 있다[1]. 특히 EDISON 플랫폼[2,3]은 웹에서 손쉽게 고성능 컴퓨팅(HPC)을 활용하는 계산 과학 시뮬레이션을 수행할 수 있는 환경을 제공해 주어, 많은 계산 과학 연구자들에게 인기 있는 플랫폼이 되어 왔다.

그러나 EDISON에서 실행되는 온라인 HPC 시뮬레이션(이하 EDISON 시뮬레이션)은 사용자에게 두 가지 불편 사항이 있다. 첫째, 실행한 시뮬레이션이 언제 끝날 지 알 수 없다. 둘째, 어떤 시뮬레이션 파라미터를 입력해야 할 지 사용에 익숙하지 않은 사용자의 경우 이를 알기 어렵다.

두 문제를 해소하기 위해 본 논문은 시뮬레이션 이력 데이터 서비스 프레임워크인 SantaFe (SimulAtion-ProveNance daTA service FramEwork)를 소개한다. 본 논문에서 제시하는 프레임워크는 웹 브라우저 환경에서 시뮬레이션 실행 시간을 추정하고 EDISON에서 제공하는 각 시뮬레이션 프로그램에 사용된 빈도 기준 상위 입력 파라미터 셋 정보를 제공한다.

SantaFe는 반복적인 시뮬레이션 프로그램 선택 화면과 입력 파라미터셋 화면 출력이 필요하기 때문에, 불필요한 중복 렌더링, 화면 전환을 피하고 코드 재사용성을 높이기 위해 React 라이브러리를 활용하여

구현되었다. SantaFe는 Node.js를 활용하여 시간 추정을 위해 사전 학습된 R 모델[4]을 실행할 수 있도록 MongoDB 질의 수행 비동기 서버를 구축하여 또한 구현되었다.

본 논문에서는 2장에서 관련 연구를, 3장에서는 설계, 4장에서는 구현 내용을 소개한다. 이 후 5장에서는 구현 최종 결론을 기술한다.

2. 관련 연구

기존 연구인 SPA 프레임워크[5]는 SantaFe와 흡사한 시뮬레이션 실행 시간 추정 서비스와 입력 파라미터 정보 서비스를 제공하기 위해 만들어졌다. 그럼에도 본 SantaFe는 SPA 프레임워크[5]의 “사용성”을 다음과 같이 크게 개선시켰다. 첫째, 기존 프레임워크에서는 극히 제한된 소수의 일부 시뮬레이션 프로그램만 지원했으나, SantaFe는 나노물리, 전산열유체, 계산화학, 전산 의학 등 네 개의 계산과학분야에서 제공한 총 19종의 시뮬레이션 프로그램에 대한 이력 데이터 서비스를 총체적으로 제공한다. 또한 동일 파라미터에 대한 실행 이력이 존재할 경우에는 실제 실행 시간 데이터의 중앙값과 추정 시간을 함께 제공함으로써 사용자에게 추정 결과가 실제 실행 시간에 얼마나 근사한지에 관해 알 수 있게끔 도움을 준다. 둘째, 오프라인에서 실행된 결과를 단순히 보여주는 기존 프레임워크 [5]와는 달리, 본 논문에서 제안하는 프레임워크는 온라인에서 주어진 입력 파라미터에 대해 R로 쓰여진 사전 학습된 시간 추정 모델을 동적으로

내부 호출하고 실행하여 그 추정 결과를 웹 브라우저 상에서 바로 제공한다. 셋째, 기존 프레임워크와는 달리 SantaFe는 다음에서 상세히 언급할 React[5]를 이용하여 재사용성 높은 사용자 인터페이스를 제공하여 효율적인 렌더링이 가능하도록 하였다.

React[5]는 Facebook 개발팀에 의해 개발된 Front-end 자바스크립트 라이브러리의 일종이다. React의 경우 SPA(Single Page Application) 방식을 사용한다. 기존 웹 사이트에서 Client의 요청에 따라 매번 새로운 페이지를 서버에서 전송해주었다. 하지만 React의 경우 하나의 페이지만을 통해 동적으로 요소들을 바꾸어 가면 효율적인 렌더링 과정을 진행한다. 또한 Virtual DOM을 통해 최소한의 변경사항만을 렌더링하여 사용자 인터페이스 전환 효율성을 개선한다. React 코드는 Component 기반 구조를 가지고 있기 때문에 코드의 부품화, 재사용성을 높일 수 있다. 본 논문의 웹 시스템의 경우 반복적인 타입 선택 페이지로 진행되기 때문에 이러한 React의 장점을 살려 렌더링의 효율을 높이고 코드의 재사용성을 높였다.

3. 설 계

3.1 시스템 구조

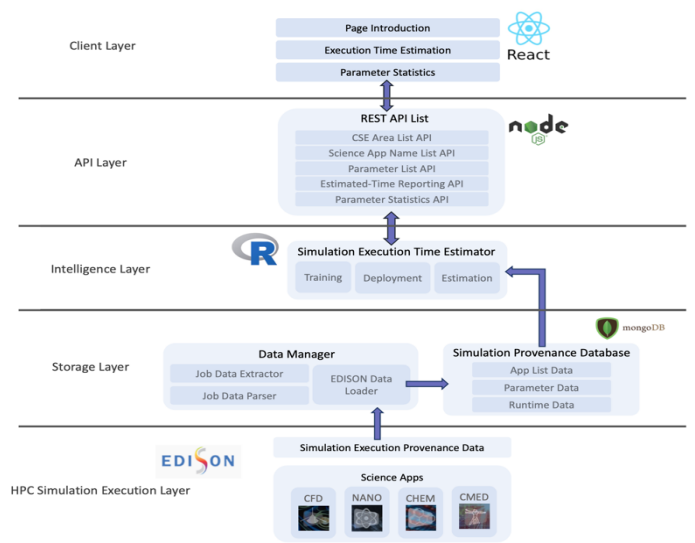


그림 1. 시스템 구조도

그림 1은 본 웹 시스템의 구조를 나타낸다. 본 시스템은 계층적 구조를 띄고 있다. 가장 아래의 HPC Simulation Execution 계층에서는 EDISON 플랫폼에서 Simulation Execution Provenance Data를 수집한다. 이후 Storage Layer에서는 얻은 데이터들을 가공 후 mongoDB에 저장한다. Intelligence Layer는 가공된 데이터들을 통해 기계학습 모델을 생성시킨다. 이후 시뮬레이션 시간 추정 서비스는 본 기계학습 모델을

통해 값을 얻는다. API Layer의 경우 크게 5개의 API 항목들을 통해 Client Layer에게 Json 형식 데이터를 제공한다. Client Layer는 React 라이브러리를 사용하여 사용자 인터페이스를 제공한다.

본 시스템의 코드는 GitHub¹에 공개되어 있다.

4. 기능

이장에서는 제안된 프레임워크의 두 가지 주요한 기능인 시뮬레이션 시간 추정 기능과 입력 파라미터 집합 정보 제공 기능을 소개한다. 더 자세한 설명은 YouTube²에 나와 있다.

4.1 시뮬레이션 시간 추정 기능

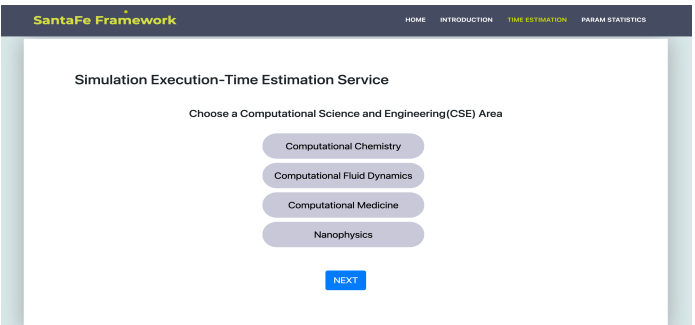


그림 2. CSE TYPE 선택 화면

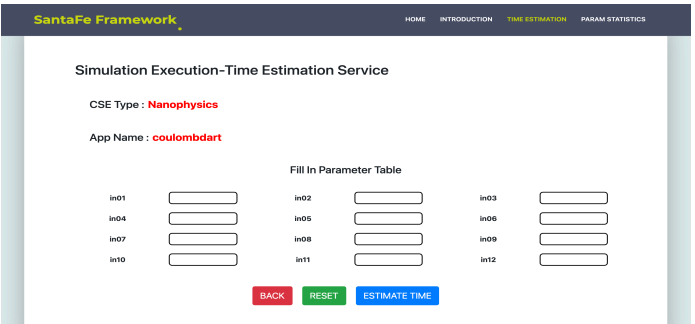


그림 3. Parameter Set 입력 화면

EDISON 플랫폼의 경우 결과를 도출하는데 까지 짧게는 3초 많게는 수만 초가 소요된다. 따라서 사용자들은 결과 도출 시간을 파악하지 못해 불편함이 유발된다. 이 불편함을 해소하기 위해 본 웹 시스템은 시뮬레이션 수행 시간을 추정할 수 있는 기능을 제공한다. 시뮬레이션 시간 추정 기능은 CSE, 시뮬레이션 프로그램명을 차례로 선택 후 추정하고 싶은 Parameter set을 입력 시 추정 시간을 확인해 볼 수 있다. 그림 2와 같이 CSE, 시뮬레이션 프로그램명을 선택할 수 있는 화면이 차례로 진행된다. 화면 전환은 페이지, URL 전환 없이 부분적인 Component만을 교체해주어 Rendering 교체를 최소화하였다. 그림 3은 선택한 CSE, 시뮬레이션 프로그램명에 해당하는 Parameter들을 입력하는 페이지이다.

¹ <https://github.com/knudeallab-papers/santafe>

² https://youtu.be/vClkeI_UIjw

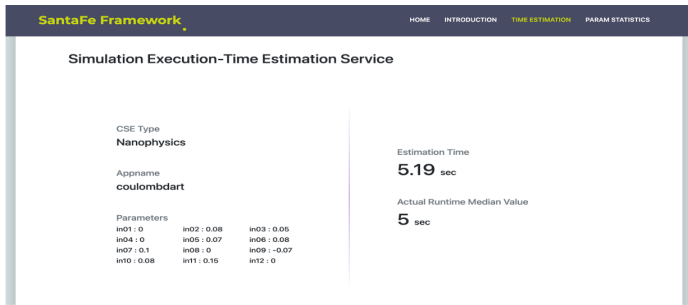


그림 4. 시뮬레이션 추정 시간 출력 화면

“BACK” 버튼을 통해 언제든 CSE, 시뮬레이션 프로그램 선택 페이지로 되돌아갈 수 있다. 최종적으로 parameter를 입력했을 시 “ESTIMATE TIME” 버튼을 통해 시뮬레이션 추정 결과를 확인해볼 수 있다.

시뮬레이션 추정 시간은 Node.js의 비동기 구문을 통해 기계학습 모델 R script를 실행하여 결과를 얻는다. 얻은 결과는 그림 4와 같이 이용자에게 전달한다. 시뮬레이션 추정 시간 이외에 mongoDB에 저장된 실제 실행시간 데이터와 동일한 Parameter set이 들어올 경우에는 실제 실행시간 데이터들 중 중앙값을 반환해 화면에 출력한다. 이를 통해 사용자들은 해당 시뮬레이션 추정 시간이 실제 실행 시간과 얼마나 근사한 값으로 추정되었는지 확인해 볼 수 있다.



그림 5. 프로그램 타입별 시뮬레이션 추정에 소요되는 시간 그래프

그림 5의 차트는 시뮬레이션 프로그램 타입 별로 시뮬레이션 추정을 위해 R script를 실행하는 소요시간을 계산한 결과이다. 소요시간은 시뮬레이션 프로그램 타입 별로 5회씩 시뮬레이션을 실행해 측정된 시간들 중 최솟값, 최댓값을 제외한 후 나머지 값들 중 중앙값을 구하는 방식으로 계산되었다. 그 결과 소요 시간은 0.02 ~ 0.06초, 평균적으로 0.03초의 짧은 시간 만에 시뮬레이션 시간을 추정하여 이용자에게 전달해 줄 수 있는 것을 확인할 수 있다. 이를 통해 이용자들은 실행한 시뮬레이션이 언제 끝날지 아주 짧은 시간 만에 확인할 수 있는 것을 보여준다.

4.2 상위 빈도 Parameter Set 확인 기능

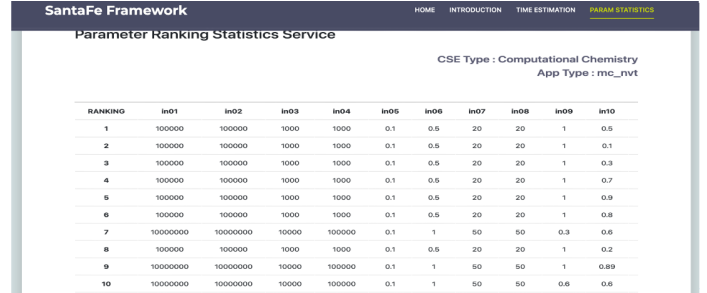


그림 5. 상위 빈도 Parameter Set 출력 화면

상위 빈도 Parameter Set 확인 화면의 경우에도 시뮬레이션 시간 추정 화면과 동일하게 CSE, 시뮬레이션 프로그램 타입을 선택하는 과정을 진행해야 한다. 따라서 시뮬레이션 시간 추정 화면에서 사용한 CSE, 시뮬레이션 프로그램 타입 선택 화면과 동일하게 진행한다. 타입을 차례로 선택하게 되면 그림 5와 같이 해당 프로그램의 시뮬레이션 Parameter Set에 대한 빈도 Top 10 을 확인해 볼 수 있다. Parameter set 통계의 경우 mongoDB에 parameter set별로 검색 횟수를 카운트하고 있기 때문에 실시간으로 검색 횟수를 업데이트하고 화면에 출력한다.

5. 결 론

본 논문은 EDISON 시뮬레이션 데이터를 활용하여 시뮬레이션 시간을 추정하고 상위 검색 순위 Parameter Set을 확인해 볼 수 있는 SantaFe 웹 시스템 개발을 기술하였다. React를 사용하여 보다 사용자 친화적이며 효율적인 렌더링을 수행하는 유저 인터페이스를 구현했으며 Node.js를 사용하여 기계학습 모델 R script, mongoDB 쿼리의 비동기적 서버 동작 과정을 구현했다.

본 논문을 통해 계산 과학 시스템에 시뮬레이션 시간을 추정해보고 parameter set을 실시간으로 확인해 보는 등 시뮬레이션 보조 웹 시스템으로써 역할하기를 기대한다.

향후 EDISON 플랫폼과 더불어 계산 과학 분야의 시뮬레이션 보조 시스템으로 활용될 수 있게끔 보다 다양한 분야의 시뮬레이션 서비스를 준비하고자 한다.

참 고 문 헌

- [1] (주)대덕넷, “계산과학, 빅데이터 과학자들 연구방식 바꿀 것”, URL: <https://www.hellodd.com/news/articleView.html?idxno=58736>, 2021년 5월 10일 확인.
- [2] Young-Kyoon Suh *et al.*, “EDISON: A Web-Based HPC Simulation Execution Framework for Large-Scale Scientific Computing Software,” in *CCGrid/16*, pp. 608–612, May 2016.
- [3] EDISON, URL: <https://www.edison.re.kr/>, 2021년 5월 10일 확인.
- [4] Young-Kyoon Suh *et al.*, “CLUTCH: A Clustering-Driven Runtime Estimation Scheme for Scientific Simulations,” *IEEE Access* 8: 220710–220722, 2020.
- [5] 정현태, 김성현, 서영균, “시뮬레이션-이력 데이터 분석 서비스 프레임워크의 설계 및 구현”, in *KCSE 2019* 제 21권 1호, pp.289–293, 2019.
- [6] React, <https://koreadjs.org/>, 2021년 5월 10일 확인.