2016 16th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing

EDISON: A Web-based HPC Simulation Execution Framework for Large-scale Scientific Computing Software

Young-Kyoon Suh, Hoon Ryu, Hangi Kim, and Kum Won Cho
Korea Institute of Science and Technology Information (KISTI),
Daejeon, Republic of Korea
{*yksuh, elec1020, hgkim, ckw*}*@kisti.re.kr*

***Abstract*—Computational science and engineering (CSE) researchers usually develop their own technology computer-aided design (TCAD) programs, accompanying large-scale computation and I/O on high-performance computing (HPC) resources like clusters or supercomputers. The researchers typically use command-line interface (CLI) such as Terminal to access the HPC resources. But CLI may not be a useful tool to those who conduct research or get educated with the TCAD software, because of their unfamiliarity with executing a series of commands. Thus there has been a strong need on a platform that assists domain-specific scientists to easily share, access, and run TCAD services. To satisfy the need, in this poster we present a novel cyber-environment, called "Education-research Integration through Simulation On the Net" (EDISON), which has been designed and implemented to access and run various TCAD software tools developed in five selected CSE fields over the past four years. The EDISON platform comprises three layers: *application portal* to browse and run TCAD software, *middleware* to manage metadata associated with TCAD software and handle online simulation jobs, and *infrastructure* to support network, storage, and computing resources. In this demo a user will interact with the EDISON platform to perform two representative use-case scenarios: 1) browsing various TCAD tools, selecting one of the tools, controlling with its parameters, and running a simulation job from the tool, and 2) constructing a scientific workflow of selected TCAD tools and executing the workflow. At the end, the user will visualize the completed simulation results.**

***Keywords*-HPC; Simulation; Cyber-infrastructure; Middleware; Scientific Workflow;**

I. Introduction

In the past years computational science and engineering (CSE) fields have become more popular, with the aid of high-performance computing (HPC) services remarkably advancing in computing power, network, and storage resources. In this wave, CSE researchers have been getting more interested in solving not only much more compute-intensive problems accompanying huge I/O operations, but also successfully demonstrating their research results that can educate CSE students in a user-friendly environment.

CSE researchers typically build their own technology computer-aided design (TCAD) softwares that run in parallel on HPC resources such as supercomputers and clusters. The researchers are familiar with command line interface (CLI) such as Terminal as a gateway to access the HPC resources. However, CLI may not be a useful tool to those who conduct research or get educated with the TCAD software services, as they are most likely not to be familiar with executing a series of commands to configure and run the TCAD programs on clusters. Hence, there has been a strong need on a TCAD-centric platform, or *Science Gateway* [1], supporting graphical user interface (GUI) such as web portals, where CSE domain-specific scientists can easily register, access, and run TCAD services, performing computation-heavy and IO-intensive online simulations on the HPC resources.

To fulfill this demand raised by the CSE community, the Korean government has launched a new project, called "EDucation-research Integration through Simulation On the Net" (EDISON), based on the underlying, strong IT infrastructure in Korea. This project is a joint work between the KISTI main center focusing on developing and managing a general hub (platform), or EDISON, of a variety of discipline-neutral TCAD softwares and a group of (five) domain-specific centers responsible for supplying the TCAD software tools to be distributed on the EDISON platform. As of now, this hub has been sufficiently robust to service 28,052 people, 249 TCAD software tools, and 365 contents over the past four years. Thanks to its technological superiority and growing popularity in Korea, our platform was successfully deployed (in June 2015) at National Institute of High-performance Computing (NCHC) in Taiwan. The EDISON platform will continue to expand its reputation by being shipped to Vietnam's Institute of Computational Science (ICS) this year. The platform is expected to make a great success in serving their CSE researchers and students.

In this poster we present our EDISON platform, consisting of three layers: *application*, *middleware*, and *infrastructure*. We describe how this platform operates in concert of these three layers: efficiently managing versatile TCAD software tools browsed on EDISON web portals at the application layer and performing simulation jobs with heavy computation and I/O submitted to KISTI's HPC resources at the infrastructure layer through the middleware layer. In the Appendix we manifest two representative demonstration scenarios of utilizing the EDISON platform for audience.

978-1-5090-2453-7/16 $31.00 © 2016 IEEE
DOI 10.1109/CCGrid.2016.31

The key contributions of this poster are as follows.

- We present a novel cyber-infrastructure, called EDISON, dedicated to assist the CSE community to easily access and run TCAD services in parallel on the HPC resources such as clusters and supercomputers.
- EDISON provides a rigorous abstraction for performing online simulation work, initiated by a variety of TCAD tools used in different CSE disciplines.
- EDISON presents core middleware, consisting of 1) *metadata management framework* to manage the information about TCAD tools registered into the platform, 2) *job management framework* to schedule and manage jobs submitted through portal to HPC resources, and 3) *workflow management framework* to author and run a scientific workflow of TCAD tools.
- EDISON provides a reusable GUI architecture. This GUI architecture is applied to different web portals of five participating CSE areas: computational fluid dynamics (CFD), computational chemistry (Chem), nano physics (Nano), computer-aided optimal design (COD), and computational structural dynamics (CSD).

II. Related Work

Several HPC service platforms have been designed to support online simulations in various fields of CSE [2], [3], [4], [5]. The platforms' user-friendly GUI has greatly contributed to easy-to-run simulations. Indeed, they are leading science gateways, providing public services for online simulations.

However, there are some limitations in the existing HPC service platforms [2], [3], [5]. First, they are dedicated for simulations focusing on a specific domain area such as chemistry [5], earth science [2], and nanoelectronics [3]. Our mission is to establish a *general-purpose*, *open* hub to accommodate many TCAD tools from a variety of CSE fields. Of course, there has been around a well-known open platform, called HUBzero [4]. But the HUBzero charges a non-trivial burden on TCAD developers. Specifically, developers have to further wrap their TCAD tools in the HUBzero-provided toolkit (called Rappture) to deploy the tools on the platform by building a separate parser to interpret XML-based input scripts for their TCAD tools. We want an "easy-to-use" platform that can take the developers' code as it is. Furthermore, it is hard to guarantee service-quality due to a long distance in spite of equipped high-speed network facilities, as all the back-end HPC resource integrated into the existing platforms are located in USA. It is of critical importance to provide little network latency in performing TCAD online simulations that typically need to handle large datasets.

To overcome these limitations, we have designed and developed our own domain-neural HPC service hub of TCAD software, not only providing high-quality online simulation services but also having easy access to CSE researchers and students in Korea and some Asia-pacific areas.

III. Architecture

In this section we discuss each component of the EDISON platform. Figure 1 illustrates the overall architecture of EDISON. The EDISON platform consists of three layers: application portal service, middleware, and infrastructure.

A. The Application Portal Service Layer

To make the open computing service available, an HPC service platform should enable i) geographically-distributed researchers to share their in-house TCAD tools and associated datasets and ii) any users to run the published tools with few constraints (such as no license purchase). The platform should also provide an easy-to-use GUI, such that the researchers and users can readily perform the sharing and running activities. For the platform to be general-purpose, it should furnish easy portability for accommodating a variety of CSE fields as well. Hence, a "web-based problem-solving" environment is considered the most suitable choice to satisfy the above conditions.

The Liferay portal framework [6] has provided a perfect fit for our needs. Based on secure single-sign-on technology, the Liferay framework supports easy portlet-based plugins to dynamically compose middleware services that can carry out user authentication, TCAD tool management, and HPC simulation execution. Furthermore, it offers performance statistics for all portlets and pages for our administrators.

The application service layer, embodied by the portal framework, is composed of a variety of portlets exposed to EDISON users. These portlets have access to the aforementioned middleware services via the platform's open APIs, implemented by a web-standard REST (REpresentational State Transfer) full interface [7]. The EDISON mobile web services are also implemented through the same open APIs.

Now we elaborate on the middleware stack of EDISON.

B. The Middleware Layer

As mentioned before, the middleware layer is responsible for user authentication and authorization, TCAD software (a.k.a. solver) metadata, simulation dataset, simulation history (provenance) management, scientific workflow execution, and heterogeneous (physical and virtual) computing resource management. This middleware layer is composed of three core elements: (i) metadata management, (ii) job management, and (iii) workflow management frameworks.

Metadata Management Framework: In general, the simulation environment is formed by various components including preprocessors (e.g., control of simulation inputs), TCAD software, postprocessors (e.g., visualization tools), computing resources, and job manager. When establishing our domain-independent cyber-environment using these components, we are faced with a challenge: "no standardized data schema and interface" to access the components. Recall that existing simulation environments [2], [3], [4], [5] have supported domain-specific TCAD software.

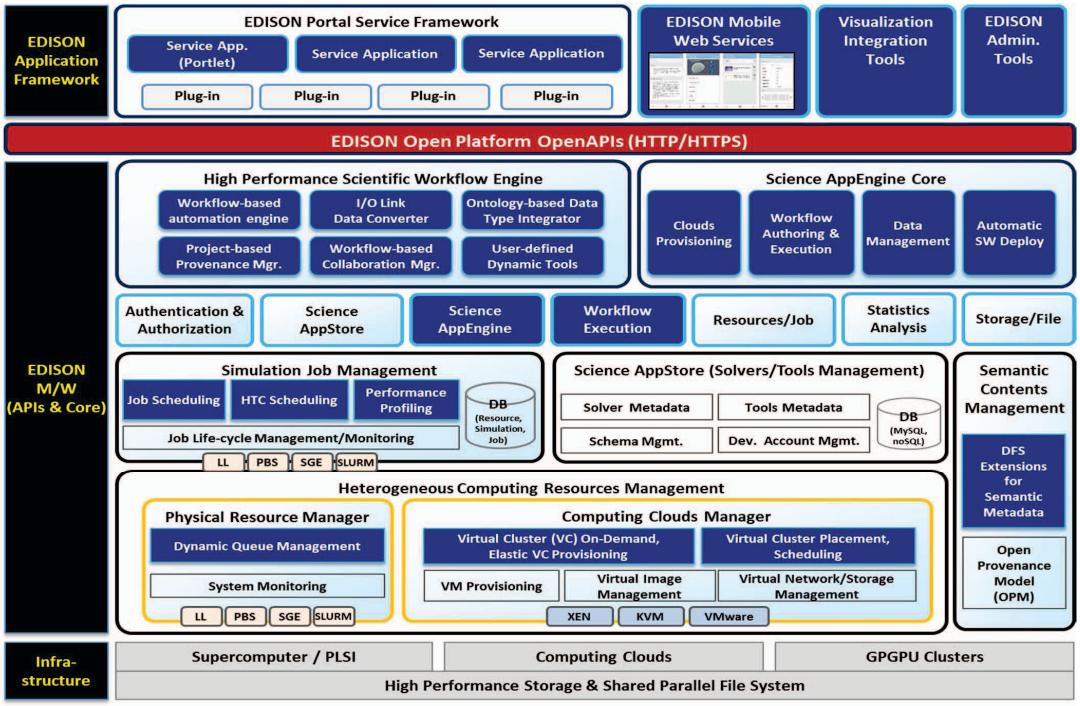


Figure 1. The Overall EDISON Architecture

To attack the challenge, we have developed the EDISON metadata management framework, called *Science AppStore*, to store and manage the metadata of a variety of TCAD software. Science AppStore is operated by a metadata manager, called *SpyGlass*. Figure 2 illustrates the architecture of Science AppStore, consisting of a metadata repository and a set of APIs to register and manage the metadata.

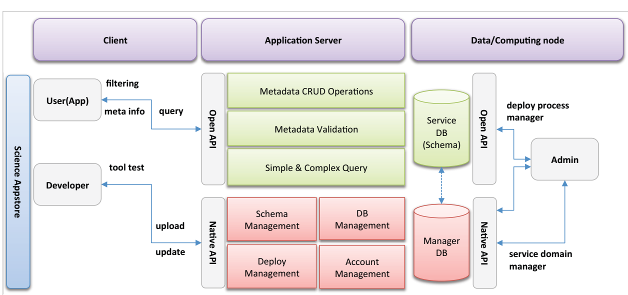


Figure 2. The Science AppStore Framework

By examining various TCAD software in the five CSE domains—CFD, Nano, Chem, COD, and CSD–described at the end of Section I, we have designed a global schema, consisting of a total of 79 elements and 53 attributes of software metadata. Considering interoperations among the above simulation environment components, we have further categorized the global schema into the following six sections. *Identification section* represents the information for TCAD software identification, including title, version, developer, and affiliation, as well as the information for EDISON users, including description, features, and screen-shot images regarding simulation detail. *Code section* represents the information for creating automated job commands, including executable file name and path and result file path, as well as the information for installing and running, including programming language, compiler, and static libraries for linker. *Parameter section* represents the information for creating an interface of simulation parameters, including the names, data types, descriptions, and default values of control parameters (that is, command-line arguments). *Category section* represents the information about simulation workflows (or, preprocessor, TCAD software, and postprocessor in sequence) and favorite TCAD software for users. *Additional section* represents the information for contents associated with TCAD software, including technical reports, research articles, lecture notes, and other contents. *System section* represents the administrator information—version control, usage frequency, and service status—of TCAD software executed on the EDISON platform.

The metadata repository has been developed using a Not-only SQL (NoSQL) [8] database system, useful to preserve unstructured (arbitrary) datasets. Specifically, we have employed MongoDB [9], allowing for the independent and expandable synchronization between simulation components with no constraints on the schema structure yet still supporting the preservation of the document-centric schema.

The APIs are exposed as a RESTFul interface and implemented with Node.js [10]. These APIs have been designed to support CRUD (create/read/update/delete) and query functions to retrieve and store the metadata of specific TCAD software, pre/post-processor, and configuration.

Virtualized Resource and Job Management Framework: The virtualized computing resource/job management framework of the EDISON platform has been designed and developed along with two design principles: 1) efficient utilization of available computing and storage resources with solid user management and 2) web-standard interface allowing system administrators to easily monitor simulation jobs and resources. The framework, managed by a job/resource manager, called *Icebreaker*, consists of three layers (in bottom-up fashion): (i) *abstraction* for inter-operations among various functional environments including user authentication, resource virtualization, and simulation job and resource management, (ii) *core-framework* for actual services for user management, (physical/virtual) server provisioning, and submitted simulation jobs, and (iii) *web-service* for RESTFul interface support for system administrators accessing and managing the virtualized resources and simulation jobs.

Workflow Management Framework: Figure 3 shows the overall architecture of the EDISON workflow framework, which allows users to build and execute a scientific workflow with different TCAD simulation tools. This framework is called SIMFLOW [11]. It consists of i) a Java-based *workbench* providing a web-based UI to author and save a workflow and ii) an *execution engine* to run the saved workflow within the workbench. The workbench teams up with SpyGlass for having access to TCAD tools' metadata and is implemented by various technologies including Spring, jQuery (UI), and jsPlumb. The engine collaborates with Icebreaker for workflow execution and is implemented by the Play Framework and Akka, supporting concurrent executions of the workflow.

C. The Infrastructure Layer

Our infrastructure supports large-scale computation and I/O initiated by simulation jobs of TCAD tools. Computing resources are comprised of supercomputer/PLSI (Partnership and Leadership for the nationwide Supercomputing Infrastructure), computing clouds (clusters) consisting of 1,168 physical cores, and GPGPU clusters. To cost-effectively absorb heavy I/O traffic, we have organized a two-tier architecture, in which we leverage i) GlusterFS installed on two high-end SSD nodes with a total of about 12 TB as "front-end" and ii) NFS on an HDD-based appliance totaling 48 TB as "back-end" in case of lack of space in the SSDs.

IV. CONCLUSION

In this poster we presented the EDISON platform for not only supporting researchers but also educating students engaged in the CSE fields. We elaborated each of the EDISON components along with the three layers. In the emerging era of convergence, we believe that EDISON will continue to expand its versatility and gain its popularity by accommodating more CSE areas involving atomic energy, biology, materials, mathematics, and pharmaceuticals.

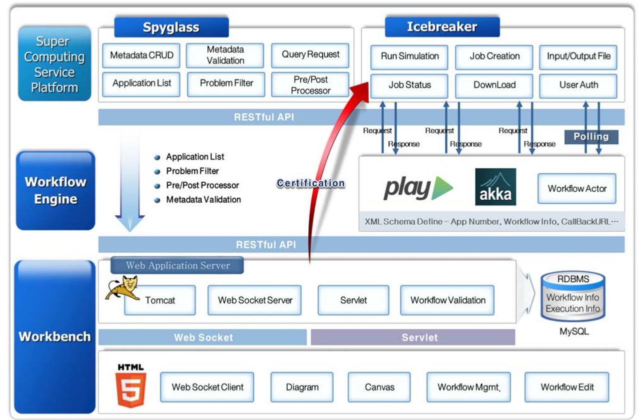


Figure 3. The EDISON Workflow Framework

ACKNOWLEDGMENT

This work was funded by the EDISON program through the National Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2011-0020576). The authors thank the EDISON team at KISTI for their efforts and anonymous reviewers for their comments.

REFERENCES

[1] H. Ryu et. al., "EDISON Science Gateway: A Cyber-Environment for Domain-Neutral Scientific Computing," *IEICE Trans.*, vol. 97-D, no. 8, pp. 1953–1964, 2014.

[2] T. J. Hacker *et al.*, "The NEEShub Cyberinfrastructure for Earthquake Engineering," *Computing in Science & Engineering*, vol. 13, no. 4, pp. 67–78, 2011.

[3] G. Klimeck *et al.*, "nanoHUB.org: Advancing Education and Research in Nanotechnology," *Computing in Science & Engineering*, vol. 10, no. 5, pp. 17–23, 2008.

[4] M. McLennan and R. Kennell, "HUBzero: a Platform for Dissemination and Collaboration in Computational Science and Engineering," *Computing in Science & Engineering*, vol. 12, no. 2, pp. 48–53, 2010.

[5] J. Schmidt and W. Polik, "WebMO Portal (Chemistry)," http://www.webmo.net, (accessed March, 2013).

[6] Liferay Inc., "Liferay Portal Framework," http://www.liferay.com, (accessed January, 2016).

[7] C. Pautasso *et al.*, "Restful Web Services vs. "Big" Web Services: Making the Right Architectural Decision," in *WWW '08*. ACM, 2008, pp. 805–814.

[8] R. Cattell, "Scalable SQL and NoSQL Data Stores," *ACM SIGMOD Record*, vol. 39, no. 4, pp. 12–27, 2011.

[9] MongoDB, Inc., "MongoDB," https://www.mongodb.org/, (accessed September, 2016).

[10] Node.js Foundation, "Node.js," https://nodejs.org, (accessed September, 2016).

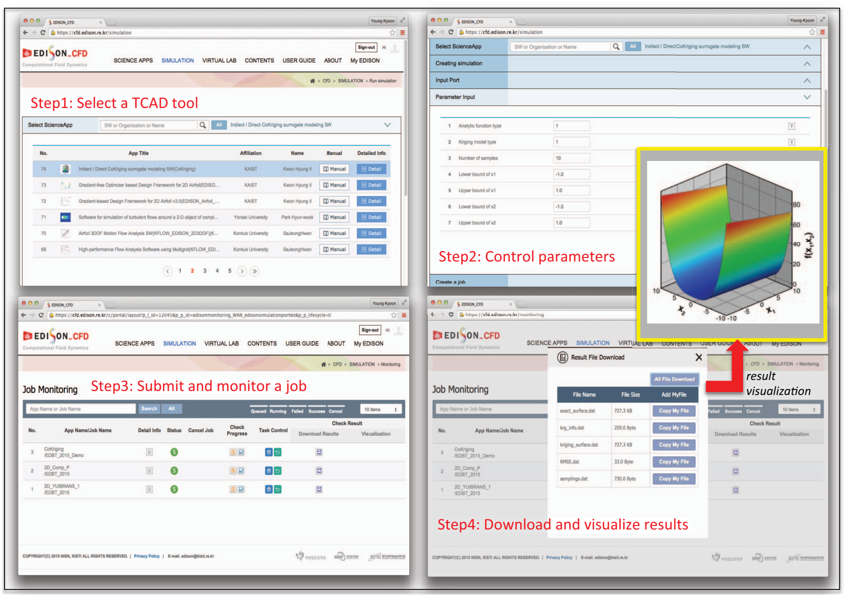
[11] H. Kim et. al., "A Design and Implementation of a Lightweight Scientific Workflow System for Higher Education and Advanced Research of Computat. Sci. and Eng." *Information*, vol. 17, no. 11, pp. 5927–5932, 2014.

Appendix
Demonstration

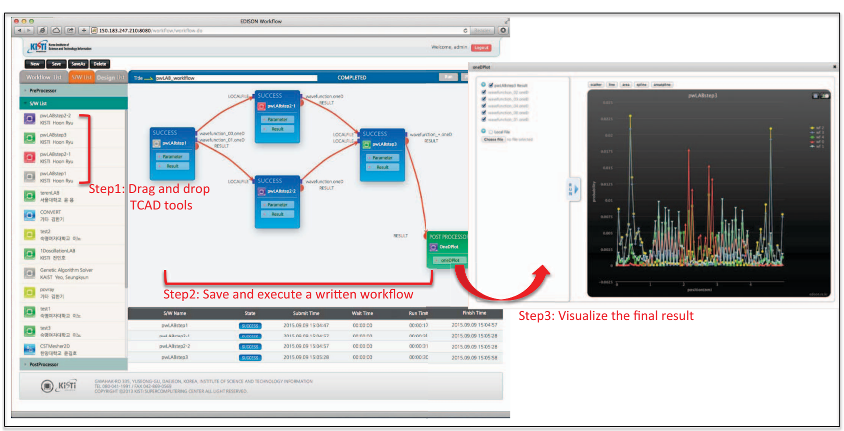
Our demonstration consists of two parts: 1) single job submission and 2) workflow authoring and execution.

The goal of the first part of this demo is to show how a user can run a TCAD tool in the EDISON system. As illustrated in Figure 4(a), here are the steps. At Step 1, an auditor has access to one of the EDISON portal sites, and she then selects one of the TCAD tools available on the chosen site. At Step 2, she controls various parameters associated with the chosen tool; she can change the default value of each parameter. At Step 3, she submits and monitors a simulation job with the controlled parameters. At Step 4, she waits for the job to be completed. Once the job is finished, she can visualize the results via an external tool.

The second part assists the user to experience a scientific workflow of connectable TCAD tools. Specifically, the user authors a workflow on the workbench, executes it through the engine, and finally carries out the visualization of the completed results of that workflow. As depicted in Figure 4(b), the steps are following. At Step 1, she logs in to our workflow engine site and browses all TCAD software tools that can be connected to each other. She then places the selected tools through drag-and-drop on the right panel. By matching input and output ports of the tools, she completes authoring the workflow. At Step 2, she saves the written workflow and subsequently runs that workflow through the engine. She can monitor how each state of the workflow components gets updated. At Step 3, after completing the execution, she can then visualize the final results by a click.



(a) Running a Simulation Job on a Selected TCAD Software



(b) Running a Scientific Workflow

Figure 4. EDISON demo scenarios