Science as a Service

Ravi Madduri and Ian Foster | Argonne National Laboratory and the University of Chicago

Researchers are increasingly taking advantage of advances in cloud computing to make data analysis available as a service. As we see from the articles in this special issue, the science-as-a-service approach has many advantages: it accelerates the discovery process via a separation of concerns, with computational experts creating, managing, and improving services, and researchers using them for scientific discovery. We also see that making scientific software available as a service can lower costs and pave the way for sustainable scientific software. In addition, science services let users share their analyses, discover what others have done, and provide infrastructure for reproducing results, reanalyzing data, backward tracking rare or interesting events, performing uncertainty analysis, and verifying and validating experiments. Generally speaking, this approach lowers barriers to entry to large-scale analysis for theorists, students, and nonexperts in high-performance computing. It permits rapid hypothesis testing and exploration as well as serving as a valuable tool for teaching.

Computation and automation are vital in many scientific domains. For example, the decreased sequencing costs in biology have transformed the field from a data-limited to a computationally-limited discipline. Increasingly, researchers must process hundreds of sequenced genomes to determine statistical significance of variants. When datasets were small, they could be analyzed on PCs in modest amounts of time: a few hours or perhaps overnight. However, this approach does not scale to large, next-generation sequencing datasets—instead, researchers require high-performance computers and parallel
algorithms if they are to analyze their data in a timely manner. By leveraging services such as the cloud-based Globus Genomics, researchers can analyze hundreds of genomes in parallel using just a browser.

In this special issue, we present three great examples of efforts in science as a service. In “A Case for Data Commons: Toward Data Science as a Service,” Robert L. Grossman and his colleagues present a flexible computational infrastructure that supports various activities in the data life cycle such as discovery, storage, analysis, and long-term archiving. The authors present a vision to create a data commons and discuss challenges that result from a lack of appropriate standards.

In “MRICloud: Delivering High-Throughput MRI Neuroinformatics as Cloud-Based Software as a Service,” Susumu Mori and colleagues present MRICloud, a science as a service for large-scale analysis of brain images. This article illustrates how researchers can make novel analysis capabilities available to the scientific community at large by outsourcing key capabilities such as high-performance computing.

Finally, in “WaveformECG: A Platform for Visualizing, Annotating, and Analyzing ECG Data,” Raymond Winslow and colleagues present a service for analyzing electrocardiogram data that lets researchers upload time-series ECG data and provides analysis capabilities to enable discovery of the underlying aspects of heart disease. WaveformECG is accessible through a browser and provides interactive analysis, visualization, and annotation of waveforms using standard medical terminology.

As adoption of public cloud computing resources for science increases, science as a service provides a great way to create sustainable, reliable services that accelerate the scientific discovery process and improve the adoption of various tools and thus increase software reuse.

Ravi Madduri is a project manager and a Senior Fellow at the Computation Institute at the Argonne National Laboratory and the University of Chicago. His research interests include high-performance computing, workflow technologies, and distributed computing. Madduri has an MS in computer science from the Illinois Institute of Technology. Contact him at rm@anl.gov.

Ian Foster is director of the Computation Institute, a joint institute of the University of Chicago and Argonne National Laboratory. He is also an Argonne Senior Scientist and Distinguished Fellow and the Arthur Holly Compton Distinguished Service Professor of Computer Science. His research deals with distributed, parallel, and data-intensive computing technologies, and innovative applications of those technologies to scientific problems in such domains as climate change and biomedicine. Foster received a PhD in computer science from Imperial College, United Kingdom.

Selected articles and columns from IEEE Computer Society publications are also available for free at http://ComputingNow.computer.org.

COMPUTER ENTREPRENEUR AWARD

In 1982, on the occasion of its thirtieth anniversary, the IEEE Computer Society established the Computer Entrepreneur Award to recognize and honor the technical managers and entrepreneurial leaders who are responsible for the growth of some segment of the computer industry. The efforts must have taken place over fifteen years earlier, and the industry effects must be generally and openly visible.

All members of the profession are invited to nominate a colleague who they consider most eligible to be considered for this award. Awarded to individuals whose entrepreneurial leadership is responsible for the growth of some segment of the computer industry.

DEADLINE FOR 2017 AWARD NOMINATIONS
DUE: 15 OCTOBER 2016

AWARD SITE: https://www.computer.org/web/awards/entrepreneur
www.computer.org/awards