



Echo software revolutionizes data analysis

March 16, 2017

New LANL software offers data analysts tremendous gains in accuracy, efficiency, and transparency

Online shopping tools help customers locate hard-to-find products without visiting multiple stores or quickly compare similar products across a wide range of sellers. Shoppers can see detailed specifications at a level of detail not typically included on product packaging, and they can find detailed reviews of products to learn where they were manufactured, what they are composed of, or what the differences are in the latest version of a given product. All of this information enables modern shoppers to make decisions more accurately, more efficiently, and with a greater understanding of their ultimate selection.

The myriad of products available to modern shoppers pales in comparison to the mountains of data being generated every day. Analyzing data piecemeal is usually uninteresting and uninformative. Analysts need tools to evaluate multiples pieces of data simultaneously that are related by a common thread. Identifying that common thread and retrieving the relevant data can be tedious and time-consuming. To identify the data needed for a given analysis, traditional methods begin with a process not unlike wandering through many physical stores. Analysts search in hard drives, share drives, and thumb drives, poring over data in spreadsheets, documents, and log files. They search for data related by certain common characteristics in their metadata (extra information about the data itself). Once the datasets have been discovered and collected, the analyst must spend time working with each dataset. The analyst converts the datasets to a common form that will permit them to be processed together, meanwhile investing significant time into ensuring that their content and metadata are believable and, ultimately, trustworthy.

At Los Alamos National Laboratory, a team of engineers in the Applied Engineering and Technology Division, Weapon Stockpile Modernization Division and the National Security Education Center has developed Echo, a new data analysis environment. Echo packages data with metadata, removing several categories of opportunity for human error while speeding and simplifying the analysis required for finding, and extracting information from, large, complex datasets. One can compare Echo's approach to data analysis with the online method of shopping – exploring the data and metadata at the same time to achieve better results. In the shopping analogy, data could be a coat; metadata would be the color, size, cut, closures, etc. In Echo, metadata helps users

find data, manage and organize analysis workflows, and accurately report results as relevant, usable information.

Traditional analysis software focuses on just the data; the metadata is left in the hands of the analyst. Results are produced using scripts that the analyst uses to combine data in different ways and perform mathematical operations on the data. While the analyst doing the work may have a good handle on the process, many of the decisions and other contextual information that led to the result are only present in the analyst's mind. How the result of the analysis is produced is invisible to other analysts or the users (customers) of the result.

Conversely, Echo keeps track of every change made, who made it, and archives each intermediate step in an analysis. The graphic shows the path that one analysis has taken. At the bottom left is the original data. A change is made (indicated by the triangle) that assigns units, and the result is indicated by the open square. An additional operation is performed to select a subset of the data within a time window producing the next result indicated by the shaded square. The process continues until the analyst produces the desired final result. The history of each step along the way is automatically captured and any intermediate result can be reviewed and visualized. Other analysts can modify individual steps in an analysis workflow to produce new results. This capability enables collaboration among different parties, allowing them to share results with transparency and compare different analysis approaches.

Curt Ammerman, a program manager and benefactor of the software observed: "Echo was initially developed in support of a LANL project to manage and facilitate analysis of a variety of environmental and qualification test data. The Echo toolset provides our test engineers with a logical, efficient process for handling large, complex data sets on the front end that saves considerable time on the back end. In light of the substantial investments to acquire system test data, Echo maximizes the utility of these data by permitting rapid data interrogation and comparison, supporting swift generation of test specifications, streamlining figure generation and report writing, and ultimately enabling informed decisions. Echo has paid for itself many times."

With Echo, the data and metadata are encapsulated, reducing the potential for error and improving the access to all of the information required to perform an analysis and document the results. Echo was originally developed in support of systems qualification for the weapons program; however, the software is an incredibly powerful tool that can be used in a wide variety of applications. Scientists and data analysts can use the common framework of Echo to manage and analyze many types of datasets including market data in the field of economics and finance, diagnostic imagery in the area of medical science, and geospatial measurements used in climate science. The utility of this analysis management program is only beginning to be realized. As more users recognize its power, it will be applied in many more applications.

Los Alamos National Laboratory

www.lanl.gov

(505) 667-7000

Los Alamos, NM

Managed by Triad National Security, LLC for the U.S Department of Energy's NNSA

