Avoid Drowning in a Deluge of Simulation Data: The Case for Simulation Analytics

CIMdata Commentary

Key takeaways:

- *The increasing volume of simulation data is overwhelming analysts' capacity to find the useful information contained within it.*
- *Simulation analytics refers to both the process and the tools for extracting useful information from large amounts of simulation data.*
- *Simulation analytics can be applied to enrich information and reduce the volume of data, to compute simulation metrics, and to enable important data to be retained and archived.*
- *CIMdata is tracking the development and adoption of several promising new Simulation Analytics applications; we consider Tecplot Chorus to be noteworthy.*

As companies move to digital product development, the role of simulation (computer-aided engineering) rises in importance. Indeed, with the shift to model-based and simulation-driven design, the role of simulation is evolving rapidly. Leading users now leverage simulation to help create product concepts—to optimize designs to best meet competing or conflicting requirements and to create robust designs whose performance is less sensitive to variations in manufacturing processes or customer usage.

Simulation is no longer used solely to evaluate a single, nominal design that typically requires only a few runs of a particular simulation application. Now simulation is being used in statistics-based assessments and optimizations involving hundreds or thousands of runs. The sheer volume of data generated in these kinds of studies can overwhelm available resources. To realize the true value of simulation, innovative approaches to deal with the resulting flood of data are required.

Beyond simply the volume of data, the problem has two other important components: First, the meaningful information content in simulation output files is generally quite sparse, so the "interesting" results can be hard to find among all the data. Second, our ability to process (i.e., read) the data is improving far less quickly than our capability to generate and store it.

Technical computing capability to generate data doubles every eighteen months (following Moore's Law). Disk capacity doubles every twelve months (Kryder's Law). By contrast, data transfer rates (disk I/O speed) double only every 36 months. Considering these issues, *Simulation Analytics* is required. Simulation Analytics refers to both the process and tools that will:

- Enrich the information content (and perhaps dramatically reduce the volume of data that needs to be stored and retained).
- Compute CAE-specific metrics.
- Find or enable the user to find "interesting" features across multiple runs and projects.
- Provide the data in a form that can realistically be retained and archived.

As with other “big data” problems, in many simulations the location and nature of interesting results are not known ahead of time. In addition, simulation has unique metrics that often need to be calculated as a prelude to post-processing. Analytics can require extensive computing power, and data visualization is required to present and aid in the comprehension of results.

One of the most challenging disciplines for which to interpret results is fluid flow, or computational fluid dynamics (CFD). Small changes in conditions or geometry can cause large changes in the flow. The relevant interesting features of the solution occur in the interior of the domain. (By contrast, for solid mechanics, the solution extremes almost always occur on the surface of the domain.) Simple metrics and standard reports may fail to reveal important flow differences between cases. At the same time, it is not possible to manually examine thousands or even hundreds of cases. Analytics is an obvious application to mitigate these issues.

Simulation analytics is not just more of what users are already performing routinely with standard post-processing. Analytics should enrich the data to increase the information content, and it should also aid in identifying salient artifacts not only in a single result, but also in the differences across hundreds of results. By extracting and retaining the most relevant information, analytics can dramatically reduce the amount of data to be retained.

A number of applications and technologies on the market today demonstrate parts of CIMdata’s vision for simulation analytics. Some data managers extract key results as metadata. Some utilities compress large files. Some post-processors are able to generate standard reports for different simulation cases, for example, process automators that orchestrate DOE studies for robust design can display parameter trades in terms of gross or scalar results metrics.

While there are a few applications and technologies on the market today that enable parts of CIMdata’s vision for simulation analytics, there is one product, Tecplot Chorus, that provides an analytics framework for CFD that fulfills much of the vision described herein. Tecplot Chorus is from Tecplot, Inc., a Bellevue, Washington company that focuses on delivering visualization software for engineers and scientists to analyze, discover, and communicate results. In many ways, Tecplot Chorus is a result of Tecplot’s realization that often the most interesting and useful information is lost within the reams of data that are expelled during a simulation analysis exercise. Fundamentally, Tecplot Chorus provides a framework for engineers to evaluate the flow physics across systems of results with ties to the integrated quantities that describe the gross physical behavior. By evaluating results in context, engineers can identify anomalies in their results quickly, allowing them to make more informed decisions earlier in the design process.

Industrial companies have made substantial investments in improving their simulation capability, and are now able to generate immense amounts of data in statistics-based studies of product performance. Analytics tools to deal with this data deluge are required. Tecplot Chorus is an example of an application that deals with the data while acknowledging (and retaining) the complexity of the information. If the information is not extracted and understood, it is wasted. At worst, incorrect conclusions based on simple metadata metrics may lead to rework and delays, if not to product performance failures and increased liability risks. The latter is especially important if the relevant data and correct solutions to guide decision making were all present from the beginning, but got lost or drowned out in the flood of other data that would be recovered with a forensic failure analysis by either engineering or legal teams.

About CIMdata

CIMdata, an independent worldwide firm, provides strategic management consulting to maximize an enterprise's ability to design and deliver innovative products and services through the application of Product Lifecycle Management (PLM). CIMdata provides world-class knowledge, expertise, and best-practice methods on PLM. CIMdata also offers research, subscription services, publications, and education through international conferences. To learn more about CIMdata's services, visit our website at http://www.CIMdata.com or contact CIMdata at: 3909 Research Park Drive, Ann Arbor, MI 48108, USA. Tel: +1 734.668.9922. Fax: +1 734.668.1957; or at Oogststraat 20, 6004 CV Weert, The Netherlands. Tel: +31 (0) 495.533.666.