NCSU CSC Laboratory for Science of Technologies for End-to-End Enablement of Data (STEED)

Mission/Vision:
The NCSU CSC Laboratory for the Science of Technologies for End-to-End Enablement of Data (STEED) will lead national and international efforts of unlocking the value of big data by providing the tools, science and talent for next-generation technologies and infrastructure. We envision the STEED group as the leading enabler of these new methods in order to accelerate data-driven decisions.

Value:
The STEED laboratory envisions a future where technologies and tools for enabling data processing and analytics will provide significant advantages to an organization wishing to unlock the value of decisions based on the data.

Approach:
We provide technologies and tools for bridging the time gap between the acquisition of data, on the one hand, and real-time and long-term strategic and tactical decision making. We enable opening of new application domains through innovative methods in structuring and transforming data.

Benefits:
STEED
- Unlocks use of cutting-edge tools and approaches to significantly lower the barriers of entry into advanced analytics
- Provides access to world-class data-science talent
- Connects to advanced technologies and tools for easier, more efficient, and more meaningful data fusion and data readiness
- Enables scalable and flexible decision support

“WE BRIDGE THE GAP BETWEEN THE DATA AND DECISIONS AND OPEN NEW APPLICATION DOMAINS, BY STRUCTURING AND TRANSFORMING THE DATA TO UNLOCK ITS POTENTIAL.”
Center Director: Rada Chirkova  
Associate Professor of Computer Science  
chirkova@csc.ncsu.edu

Technical Director: John Streck  
Director of Institute for NEXT Generation IT Systems  
jpstrech@csc.ncsu.edu

Sample Projects

- **Formalizing data readiness under changing technologies:** Jon Doyle, SAS Institute Distinguished Professor, NCSU Computer Science.
  This project centers on formalization and evaluation of metrics and techniques for assessing data readiness across a broad range of applications. The goal of this project is to discover how properties of big data are transformed as they advance through processing steps, and to identify and model those properties that are conducive to unlocking the decision value of the data.

- **Visualization for shape and data analysis of time-varying volume ensembles:** Christopher G. Healey, Professor, NCSU Computer Science.
  This project proposes a novel combination of data and visual analytics techniques to visualize 3D, temporal, multivariate ensembles. The main goal is to allow analysts to gain insights into very large, complex, time-varying ensemble datasets. The proposed project will develop: (1) techniques to spatially decompose volumes into hierarchical structures that support shape comparison; (2) methods to visualize multiple ensemble members in ways that highlight differences in shape and data values across the members; and (3) algorithms to cluster volumes in a time-varying ensemble, and to identify common "change in shape" patterns that may be of interest to the analysts. Each of our techniques will be designed to scale smoothly by trading off individual detail for summarized overview information as more data is added to a visualization.

Founding Members

- **Storage Convergence for Map-Reduce:** Vincent W. Freeh, Associate Professor, NCSU Computer Science.
  The map-reduce programming model, best typified by Apache Hadoop, is becoming the de facto parallel data processing paradigm because of its ability to scale and tolerate failures. However, data in Hadoop can only be accessed from within a map-reduce program executing in the cluster. This isolation of data is problematic because it restricts (immediate) use of the results of map-reduce by other programs and it removes the data from the normal life cycle operations (such as backup and snapshot). We propose Storage Convergence between Hadoop clusters and enterprise-level storage systems that provides map-reduce applications efficient access to data in situ on enterprise storage.

- **Scalable data fusion:** Alyson Wilson, Associate Professor, NCSU Statistics.
  Data fusion integrates multiple information sources about the same entity into a consistent and useful representation. These techniques have shown themselves to be useful in applications from developing a common air picture for the military to assessing reliability during new product development to predicting student retention. The goal of this project is to develop fundamental tenets of scalable data fusion; the outcomes are expected to be applicable and useful in a variety of real-life application domains.

Contact Us
Rada Chirkova | Associate Professor | 919.513.3506 | chirkova@csc.ncsu.edu  
www.steed.ncsu.edu | 2276 EB2 | Campus Box 8206 | Raleigh, NC 27695