



Award #: 1835892

CSSI Element: Empowering Data-driven Discovery with a Provenance Collection, Management, and Analysis Software Infrastructure

PI: Yong Chen, Co-PIs: Dong Dai, Brian Ancell, William Hase

Institutions: Texas Tech University

- **Objective:** To design and develop a provenance collection, management, and analysis software infrastructure for high performance computing (HPC) systems and to train students
- **What is provenance?** Refers to entities, such as users, jobs, files, etc., and relationships among them
- **What are use cases?** Describing the history of data, identifying sources, parameters, or assumptions behind a given result, auditing data history and usage, and monitoring the system
- **Who are the users?** Users include HPC system administrators, domain scientists, and data scientists such as providing visual and predictive analysis
- **What is the current progress?** Have developed a component to collect users/jobs from UGE job scheduler (working on Slurm), have developed a component to leverage jobstats and changelogs to gather file system provenance, have used for preliminary visual analytics

