

In-Memory Database Design and Modeling to Support Guided and Self-Service Visual Discovery in Big-Data Context: An Autism Spectrum Disorder (ASD) Application Case

Researcher: Sean Howell
Advisor: Dr. Bih-Ru Lea
Business and Information Technology
Missouri University of Science & Technology, Rolla, MO 65409

Abstract

This research identifies how in-memory technology can be leveraged to understand big data in a case study using Autism Spectrum Disorder (ASD) data from Simons Foundation Autism Research Initiative (SFARI) database. Literature review was first conducted to compare and contrast different In-memory database design architectures, to define the role of in-memory computing in Big-Data Analytics, and to address the role of different types of Self-Service Business Intelligence (BI) in visual discovery. A data warehouse schema and related data modeling framework (i.e., attribute views, analytical views, and calculation views) were proposed and implemented to support guided- and self-service visual discovery at the data discovery stage utilizing SAP HANA In-memory Appliance. Finally, experiments were conducted to test the proposed models and framework with users ranging from novice to computer savvy. The data were analyzed and recommendations were provided for future research improvement.

Reflections on the Learning Experience

Research in the field of Information Technology is done by reviewing the current research in the field. A topic chosen is based on current or emerging technologies. The goal of this research topic is to evolve on current research findings or to add new research to the field. After selecting a topic, literature review is conducted. The literature review process will provide you with information on your topic of research (history, components, resources, etc.) and current research findings on your topic. The next step is to collect data and then define research questions. Statistical analysis is to be performed on the data sets to test the research questions.

I was already very aware of the resources available to me from my OURE project last year. This allowed me to start gathering good sources much faster. I didn't have to learn which database or journals to go to. My main focus was to utilize the databases and journals that I found to be the best and make sure that I was getting peer-reviewed sources. The best way I found to use the resources was to find a good peer-reviewed article that is relevant to my research and go through its references to find other articles suitable to your research.

I gained a good amount of knowledge on the experimental design process during this project. I was able to create a teaching case to use in an experiment and then conduct that experiment in multiple sessions. From there, the process was similar to what I did last year; testing research question to form hypotheses and conducting more analysis to see if each hypothesis can be proven. I was able to understand this last process much more in-depth than last year.

Interpreting statistical analysis and proving hypotheses involves statistical knowledge and interpretation. I was able to use the knowledge I gained from my OURE last year on different statistical methods and models to interpret the results of my research.