Thesis Portfolio

Analysis of Database Algorithms for Performance and Access Optimization

(Technical Report)

A Study of Security and Privacy Issues and Disagreements Concerning Cloud Adoption and Efficient Use

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Andrei Stan

Spring, 2021

Department of Computer Science

Table of Contents

Sociotechnical Synthesis:

Transition of Database Technology to Distributed Systems in a Security and Privacy Context

Technical Report Title:

Analysis of Database Algorithms for Performance and Access Optimization

STS Thesis:

A Study of Security and Privacy Issues and Disagreements Concerning Cloud Adoption and Efficient Use

Thesis Prospectus:

A Study of the Implications of Logistical and Social Issues in the Security and Privacy of Cloud Computing as they Relate to Public, Private, and Classified Workloads

Transition of Database Technology to Distributed Systems in a Security and Privacy Context

Optimizing computer usage has always been concerned with improving algorithmic and operational efficiency. This is especially true in databases, where the time to access stored data is particularly critical to everyday business efficiency. These algorithms focus around many areas. Indexing, querying, and big data are the specific targets in this technical report.

In the context of big data, the main focus is on distributed systems. Key benefits come from such a system, namely more efficient use of resources, particularly monetarily, by repurposing large quantities of low cost machines to parallelize execution of large tasks that would be difficult for even the most expensive, high end computing systems to handle. The types of datasets used in this scenario are larger than the memory capacity of the systems and techniques are used to partition the data, parallelize execution and processing on it, and recombine the results. The data itself may also be stored in a different way, deviating from centralized database implementations through organization free of schemas. This also allows for scaling and redundancy of data, making the distributed systems more reliable and durable than centralized systems.

In order to enable this kind of shift in processing, particular changes have to be made to not only ensure consistency of data and usability of distributed systems as if they were a single entity, but also to maintain the level of data security and privacy that a centralized system can provide. Algorithms must be put in place to handle data differently, doing additional operations on meta-data to guarantee not only consistency or intended access patterns on the machine but also through the networks connecting the distributed machines together.

Shifting to the topic of the STS Thesis, these security and privacy issues raise concerns among users, particularly through the evolution of such systems into cloud computing. Providers of cloud services offer all the benefits of distributed systems, applied in a user-friendly way. The users, however, have to shift from handling all of the security internally to allowing some of this responsibility to fll on the providers. This is particularly problematic when considering that this shift in responsibility also means a loss of some level of control of privacy and security that the user has on a centralized system. These concerns lead to a need for more than just algorithms but a discussion of the need for proper definitions of concepts, service agreements, models of responsibility and control, and identification of requirements and expectations.

The algorithmic side of this is discussed in the Technical Report while the social aspects around privacy and security are unpacked and analyzed in the STS Thesis.