Thesis Project Portfolio

Data Mesh: Practices from a Single Domain Perspective

(Technical Report)

The Impact of Cloud Computing as a Socio-Technical Phenomenon

(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

Brian Mbogo

Spring 2023

Department of Computer Science

Table of Contents

Socio-technical Synthesis

In this project I focus on cloud computing, the capability for people to request services like storage and computation from powerful, physically remote computers, which is a major part of modern personal life and business practices. The movement of computing resource ownership from the people who use them to the cloud has profound social and technical implications. I will use the actor-network theory analysis framework to address the social dimension of cloud computing proliferation. Although cloud computing is a socio-technically rich area of study, I aim only to describe how it changes existing relations between organizations, technologists, and everyday people. Thus, the use of knowledge from specific niches within cloud computing does not provide a comprehensive exploration of those niches. The questions addressed by this thesis are important because they focus on a specific development in the computing world which has wide-ranging applications, and thus, implications. Cloud computing has been wholeheartedly embraced by individuals and institutions, yet not much attention is paid to the scope of the potential social ramifications of its widespread adoption. By exploring the overall change in relations caused by cloud computing, my project will be able to pick apart both the positives and negatives. As modern businesses' data usage grows, so does their need for methods and technologies that preserve their agility for meeting business objectives in the face of the challenges posed by data. Pluralsight, a technology up-skilling and education platform uses a data mesh architecture to meet these modern data challenges. It ensures that each data endpoint team has the necessary data for its work serving independently defined internal and external customers and reduces duplication of effort when multiple teams are working with the same data. The management of the data mesh depends on established cloud infrastructure that enables people to work and build together in ways that would not be possible otherwise. The analytics

team in which I was embedded works with two internal teams and a single idea of a "large" external customer. The team's preferred technology stack includes Apache Spark, cloud analytics infrastructure, a continuous integration/deployment platform and the necessary interface for company-wide data transfer. Within this team, the result of the chosen technology stack is an efficient, largely formulaic development process for any given data product. The technical focus on the analytics team's implementation of a specific cloud-dependent practice gives valuable insight on how the socio-technical development of cloud computing crystallizes into standardized practices in the industry.