SafeCollections and Stewardship on Cloud Kotta

Yadu N. Babuji, Kyle Chard, Eamon Duede, Ian Foster
Computation Institute
University of Chicago and Argonne National Laboratory
{yadunand,chard,eduede,foster}@uchicago.edu

***Abstract*—To address these needs we present CLOUD KOTTA, a cloud-based architecture for the secure management and analysis of social science data. CLOUD KOTTA leverages reliable, secure, and scalable cloud resources to deliver capabilities to users, and removes the need for users to manage complicated infrastructure.CLOUD KOTTA implements automated, cost-aware models for efficiently provisioning tiered storage and automatically scaled compute resources.CLOUD KOTTA has been used in production for several months and currently manages approximately 10TB of data and has been used to process more than 5TB of data with over 75,000 CPU hours. It has been used for a broad variety of text analysis workflows, matrix factorization, and various machine learning algorithms, and more broadly, it supports fast, secure and cost-effective research.**

ACKNOWLEDGMENTS

Ack - Klab, DSaPP, Tristan