

Your EGD

Project Title Here

Michael Shell
School of Electrical and
Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0250
Email: <http://www.michaelshell.org/contact.html>

Homer Simpson
Twentieth Century Fox
Springfield, USA
Email: homer@thesimpsons.com

Abstract—The abstract goes here. Abstracts typically describe the context of the paper, the problem that is being solved, the approach at solving it, and identify major results. The paper is divided in five sections. These are not mandatory but give you a hint about what you should talk about in the paper. Page limit is 4 two-column pages. If there is a specific reason why you should need more, please let me know. If you want to provide longer descriptions of the dataset (e.g. tables with many rows) do it in another document, put that document online, and cite it here.

I. INTRODUCTION

Here you should provide the context of this work and describe the goal of the project – 1) what you want to extract from the big data set and the query and ML tasks you’re going to do 2) what measurements you will do to analyze the runtime performance and scalability of your tasks.

II. RELATED WORK

You should identify and analyze related work here. You can focus on other analysis of the same dataset or of different datasets.

III. DATASET PROFILING

Here you should describe the dataset, the RDD-based computations you did in order to ingest/profile the data, and the profiling results.

IV. QUERY AND LEARNING TASKS ON THE DATASET

Here you should explain which query/learning tasks you chose, why they make sense in the context of your dataset, and some examples of the actual code you used. You can use listings latex package with small font for printing code in latex.

A. Task description

B. Results

V. RUNTIME AND SCALABILITY ANALYSIS

Here you should describe the analysis of the runtime and scalability of your tasks. Say what you’re measuring and which experiments you are doing for scalability / speedup – e.g change number of cores/instances, increase the dataset size, etc. Then present and analyze the runtime results.

A. Methodology

B. Results

VI. CONCLUSION

The conclusion goes here.

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.