`Snark`: a distributed programming framework for nested arrays

Frank Austin Nothaft[1] and Michael Linderman[2]

[1]Department of Electrical Engineering and Computer Science, University of California, Berkeley
[2]Carl Icahn School of Medicine at Mount Sinai

Abstract

Nested arrays present a natural programming interface for expressing many different parallel algorithms. In this paper, we present the **S**calable **N**ested **AR**ray **K**it (`Snark`), a distributed framework for expressing computation on nested arrays.

1 Introduction

The nested array programming model was introduced by Blelloch [1] as a functional programming model for expressing data parallel computation.

Although several implementations of the nested vector model have been published, all of the implementations have been designed to exploit parallelism from multiple cores on a single node. `Snark` provides a distributed implementation of canonical nested array operations on top of the Apache Spark [2] in-memory map-reduce framework. Instead of using static techniques to extract parallelism from the computation, `Snark` dynamically inspects the structure of the nested arrays that are being computed upon to optimize the distribution of data and computation.

References

[1] Blelloch, G. E. *Vector models for data-parallel computing*, vol. 356. MIT press Cambridge, 1990.

[2] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., and Stoica, I. Spark: cluster computing with working sets. In *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing (HotCloud)* (2010), p. 10.