

DynaJET: Dynamic Java Efficiency Tuning

*Karl Taht, Ivan Mitic, Adam Barth, Emilio
Vecchio, Sameer Agarwal, Rajeev
Balasubramonian, Ryan Stutsman*

UUCS-20-001

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

April 16, 2020

Abstract

Modern data analytics frameworks like Apache Spark are deployed at thousands of organizations regularly running diverse jobs that process petabytes of data. This has led to engineers carefully tuning configurations for these frameworks, as small percentages gains translate to large cost and energy savings at scale. However, this is challenging as frameworks and their runtimes can have hundreds of knobs, some with non-linear impacts on performance. Hence, tuning can require extensive end-to-end experimentation. This leaves administrators with a choice of performing extra offline work, or performing experimentation on real workloads at the risk of significant interim performance degradation.

We introduce a dynamic tuning approach, *DynaJET*, which offers several key benefits over prior approaches. The foremost improvement of *DynaJET* is operation at a much finer granularity. This enables testing and tuning experimentation on real workloads without the need for submitting separate jobs. In addition, tuning decisions are made in real-time, such that configuration choices can account for both current the compute task and current machine state. We demonstrate *DynaJET*'s dynamic tuning atop the Spark compute engine using Java Virtual Machine (JVM) configuration parameters as an example. With just three JVM parameters, we find that a dynamic infrastructure can improve job completion time by up to 9.6% over a globally tuned JVM configuration, and at least 3.3% on average for an industrial data warehouse benchmark.