Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Myria Middleware: Unified Services for Hybrid Big Data Systems

CHALLENGE

A rapid evolution of big data systems have resulted in overlapping yet syntactically diverse capabilities, features, and limitations. Systems that use different data models (e.g., graph, relation, or array) have different value propositions—one system may be specialized for finding patterns in complex social networks, while another system offers state-of-the-art performance for machine learning. These systems must be used in concert for modern analytics, requiring organizations to try to integrate them at the application level, or else give up the benefits of specialization and adopt a generic "one-size-fits-all" approach by using one of these systems for all tasks, even those for which it performs poorly.

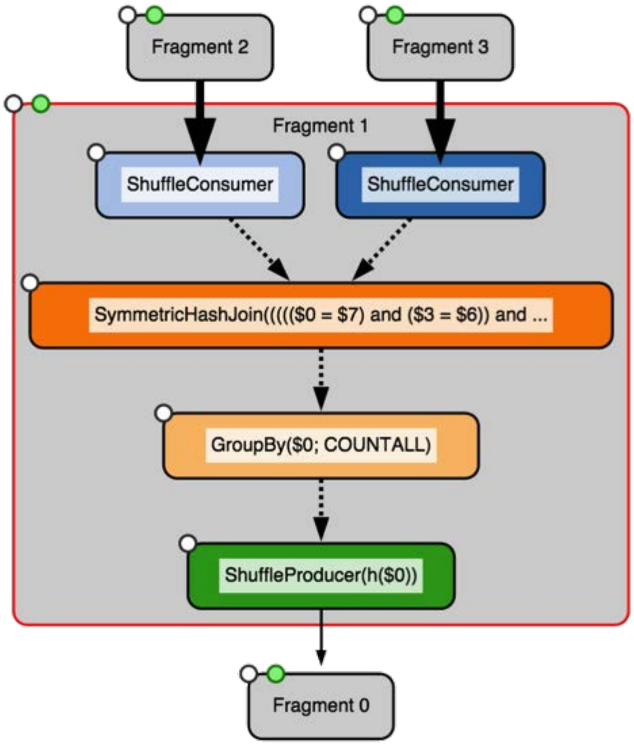
Provides monitoring and analysis services across multiple backends: plan inspection, performance profiling, per-query, and cluster utilization.

CURRENT PRACTICE

Organizations must choose either a more general "one-size-fits-all" solution and give up on opportunities for specialized solutions, or allow application maintenance costs to dominate, with each new technology requiring an enormous integration effort and significant rewrites to client applications.

TECHNICAL APPROACH

Using Myria Middleware, we aim to provide a layer of common services, languages, and programming interfaces that can insulate downstream applications from platform heterogeneity while allowing multiple



A parallel query plan in the Myria system.

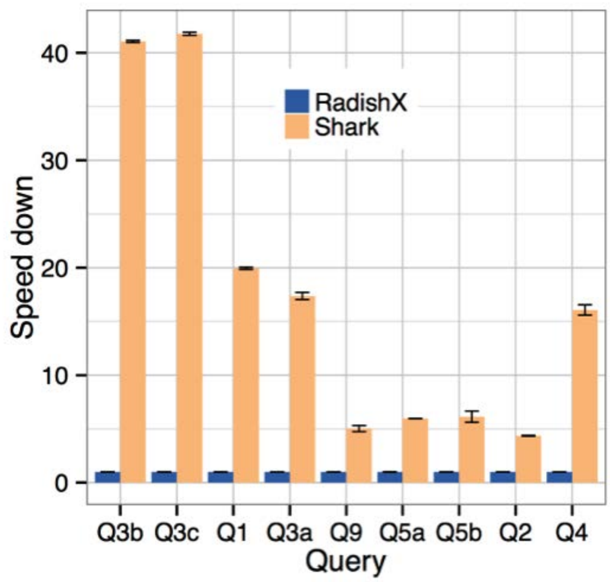
platforms—with possibly different data models—to co-exist and be used optimally on those problems for which they were designed.

Our approach has three components: a data model and programming model called the Myria Algebra, an administrative application program interface (API) for unifying administrative and system management tasks (ingest, user administration, cataloging, logging), and a workspace abstraction to provide new shared capabilities across platforms (persistent workspaces).

The insight behind our approach is that the huge diversity in the landscape of systems, algorithms, and data models for big data appears to be greater than it actually is. We find that the data and tasks supported by all of these systems typically can be described with relatively modest extensions to the relational algebra, the underlying formalism that powers relational databases. Our model captures and implements this observation to provide a unified programming model and associated services. This observation does not imply that a straightforward relational implementation can offer competitive performance, but it does provide us the ability to reason about the task, apply optimizations, and compile the resulting program to more specialized backend systems as needed.

IMPACT

With a common language runtime and the use of various backends, we have shown that Myria is competitive with state-of-the-art specialized systems for graphs, arrays, and relations—we do not find any



One of the Myria backends is far superior to Spark/Shark for graph queries. Write once, run competitively anywhere.

significant cost incurred to use a new multi-system abstraction.

With Myria, new big data systems and technology will have a direct path for integration with existing investments, significantly reducing the real costs of application integration and maintenance, as well as reducing the hidden opportunity costs of "one-size-fits-all" investments. That is, every organization wants exposure to the huge potential upside of all the activity in big data technology, but cannot be expected to accommodate the technical debt incurred by jumping from one system to another as they emerge. We are designing a platform to close this gap.

Contacts

Bill Howe
Principal Investigator
(206) 221-9261
billhowe@cs.washington.edu

John R. Johnson
Program Director
(509) 375-2651
John.Johnson@pnnl.gov

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

UNIVERSITY of WASHINGTON