Grupo de Arquitectura de Computadores
Universidade da Coruña

Performance Improvement of MapReduce Applications using Flame-MR

Jorge Veiga, Roberto R. Expósito, Guillermo L. Taboada, Juan Touriño
{jorge.veiga, rreye, taboada, juan}@udc.es

2nd NESUS Winter School & PhD Symposium 2017
Vibo Valentia, Italy
February 22th, 2017

Contents

Contents

Introduction

- Big Data has been adopted by many organizations
- Hadoop is one of the most used frameworks
- Limited performance
  - Redundant memory copies
  - Disk overhead
- Existing alternatives must rewrite applications

Introduction

- Big Data has been adopted by many organizations
- Hadoop is one of the most used frameworks
- Limited performance
  - Redundant memory copies
  - Disk overhead
- Existing alternatives must rewrite applications

Introduction

Our proposal: Flame-MR

- Completely new event-drivent architecture
- Transparent performance improvement of Hadoop applications
- In-memory computing
- Overlapping of data movement and computation

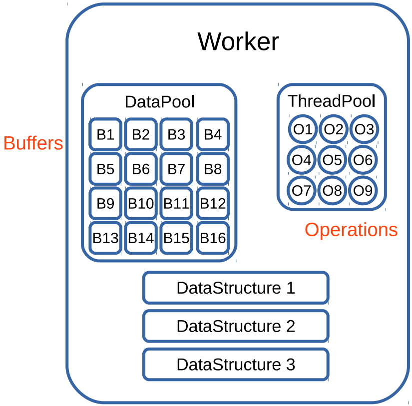
Contents

Flame-MR Design

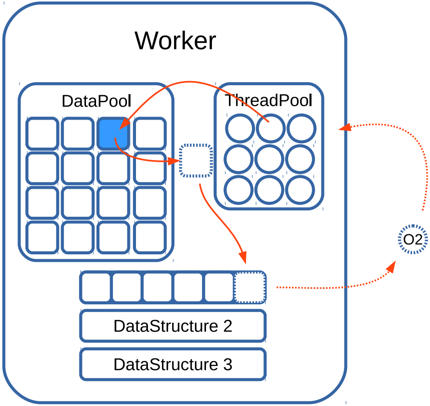
Main features

- Event-driven architecture
- Efficient memory management
- In-memory sort and merge algorithms
- Support for iterative workloads
- Full compatibility with Hadoop

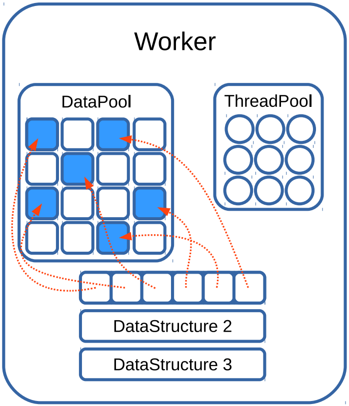
Flame-MR Design



Flame-MR Design



Flame-MR Design



Flame-MR Design

Main features

- Event-driven architecture
- Efficient memory management
- In-memory sort and merge algorithms
- Support for iterative workloads
- Full compatibility with Hadoop

Flame-MR Design

Main features

- Event-driven architecture
- Efficient memory management
- In-memory sort and merge algorithms
- Support for iterative workloads
- Full compatibility with Hadoop

Flame-MR Design

Main features

- Event-driven architecture
- Efficient memory management
- In-memory sort and merge algorithms
- Support for iterative workloads
- Full compatibility with Hadoop

Flame-MR Design

Main features

- Event-driven architecture
- Efficient memory management
- In-memory sort and merge algorithms
- Support for iterative workloads
- Full compatibility with Hadoop

Contents

Performance Results

Testbed configuration

- Evaluations conducted on Amazon EC2
  - 33 i2.4xlarge instances
  - Interconnected via GbE
  - Instance characteristics
    - 2 $\times$ 8-core Intel Xeon E5-2670 v2 2.6 GHz
    - 122 GB RAM
    - 4 $\times$ 800 GB SSD
- Experiments automated by the Big Data Evaluator tool (BDEv)
  - Configuration of the frameworks
  - Generation of input datasets
  - Collection of results
  - Available at `http://bdev.des.udc.es`

Performance Results

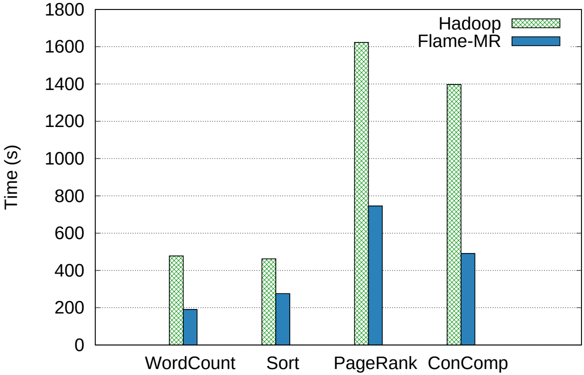
Frameworks

- Hadoop 2.7.2
- Flame-MR 1.0

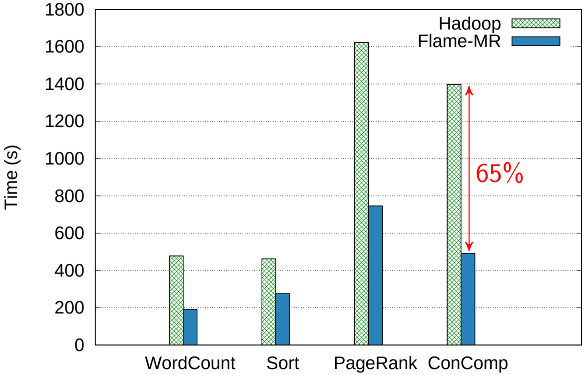
Benchmarks

Benchmark	Characterization	Input data size
WordCount	CPU bound	500 GB
Sort	I/O bound	500 GB
Connected Components	Iterative (5 iter.)	40 GB
PageRank	Iterative (5 iter.)	40 GB

Performance Results



Performance Results



Contents

Conclusions & Future Work

Conclusions

- Flame-MR improves transparently the performance of Hadoop
- Results show high performance improvements
  - Up to 65%
- Publicly available at `http://flamemr.des.udc.es`

Future work

- Development of new features
  - Automatic load balancing
- Evaluation of Flame-MR using real-world use cases

Conclusions & Future Work

Conclusions

- Flame-MR improves transparently the performance of Hadoop
- Results show high performance improvements
  - Up to 65%
- Publicly available at `http://flamemr.des.udc.es`

Future work

- Development of new features
  - Automatic load balancing
- Evaluation of Flame-MR using real-world use cases

Acknowledgments

UNIVERSIDADE DA CORUÑA

COMPUTER ARCHITECTURE GROUP
UNIVERSITY OF A CORUÑA
GAC.UDC.ES

GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD
Funded by the Ministry of Economy and Competitiveness of Spain (Projects TIN2013-42148-P and TIN2016-75845-P)

FEDER
Fondo Europeo de Desarrollo Regional
UNIÓN EUROPEA
"Una manera de hacer Europa"

Grupo de Arquitectura de Computadores
Universidade da Coruña

Performance Improvement of MapReduce Applications using Flame-MR

Jorge Veiga, Roberto R. Expósito, Guillermo L. Taboada, Juan Touriño
{jorge.veiga, rreye, taboada, juan}@udc.es

2nd NESUS Winter School & PhD Symposium 2017
Vibo Valentia, Italy
February 22th, 2017