Evaluation of a Cloud Service Provider for Distributed Data Intensive Applications

Mst. Raffael Bottoli Schemmer.

Prof. Dr. Cláudio Fernando Resin Geyer.

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

GPPD

.inf
INSTITUTO DE INFORMÁTICA
UFRGS

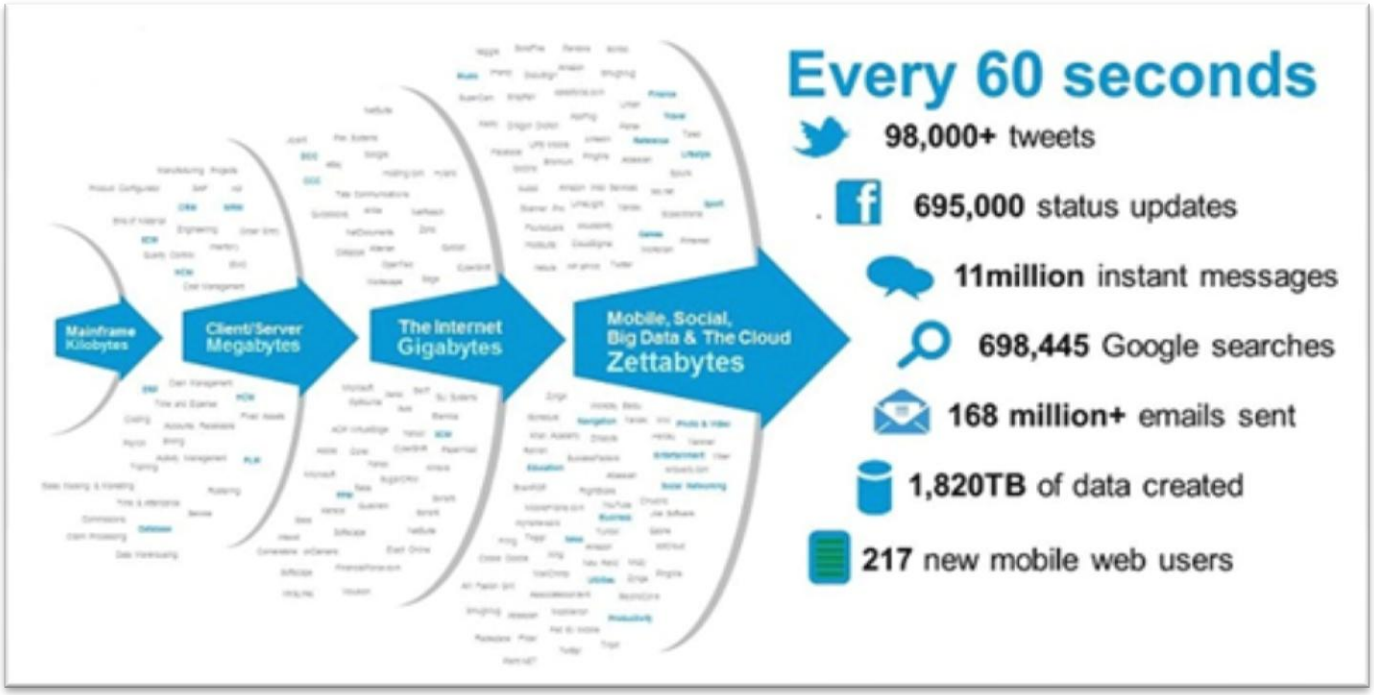
Agenda

- Introduction.
- Motivation.
- Research Goals.
- Related Concepts.
- Results.
- Conclusions.
- Future Works.

UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

.inf UFRGS

Introduction

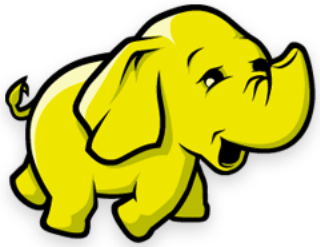


Challenges of real world. “Hp (2015)”.

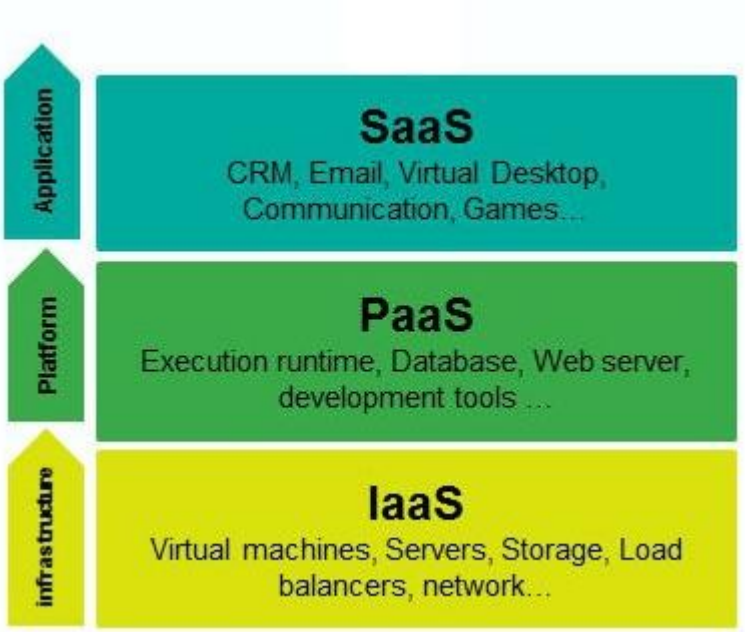
UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

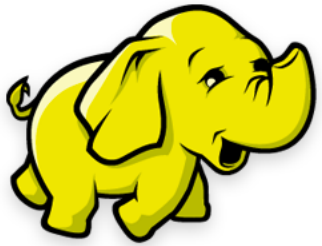
.inf UFRGS

Batch Vs. Streaming



Motivation





UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

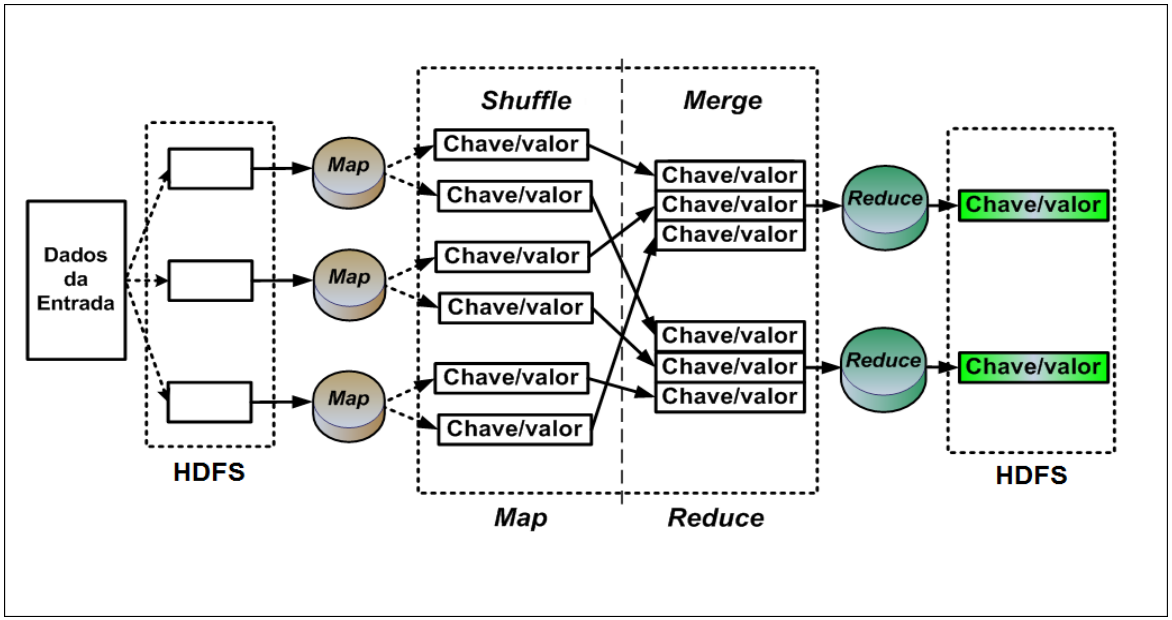
Research Goal

Evaluating Hadoop behavior running on different types of cloud computing services

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Related Concepts

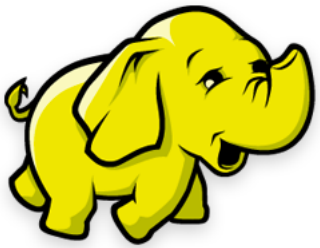


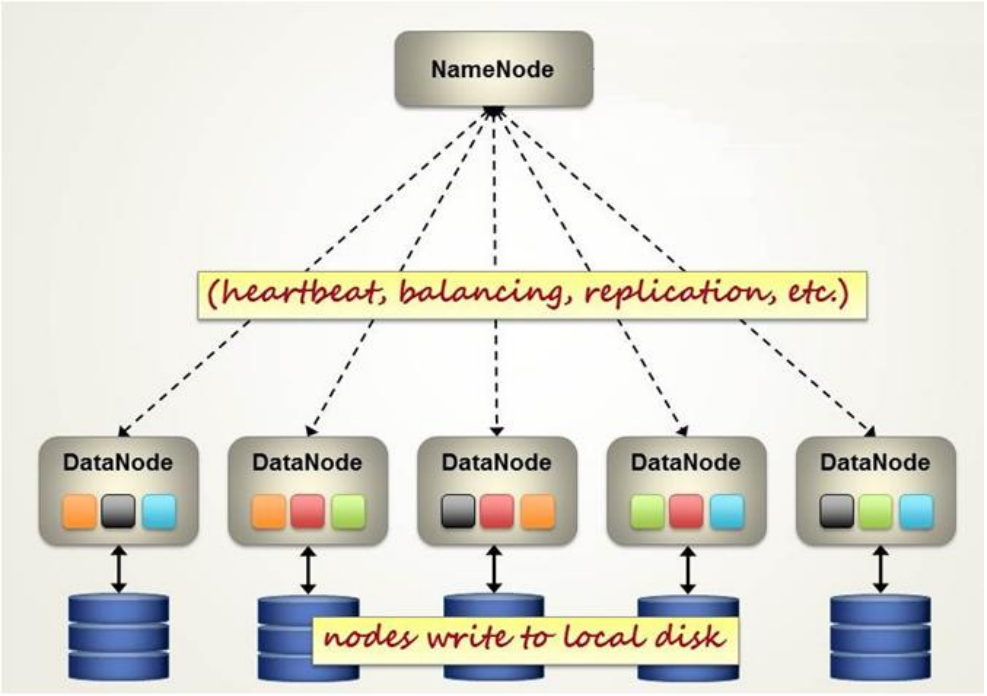
UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

hadoop MapReduce

.inf UFRGS

Related Concepts



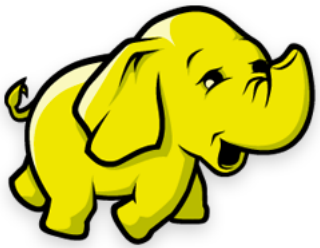


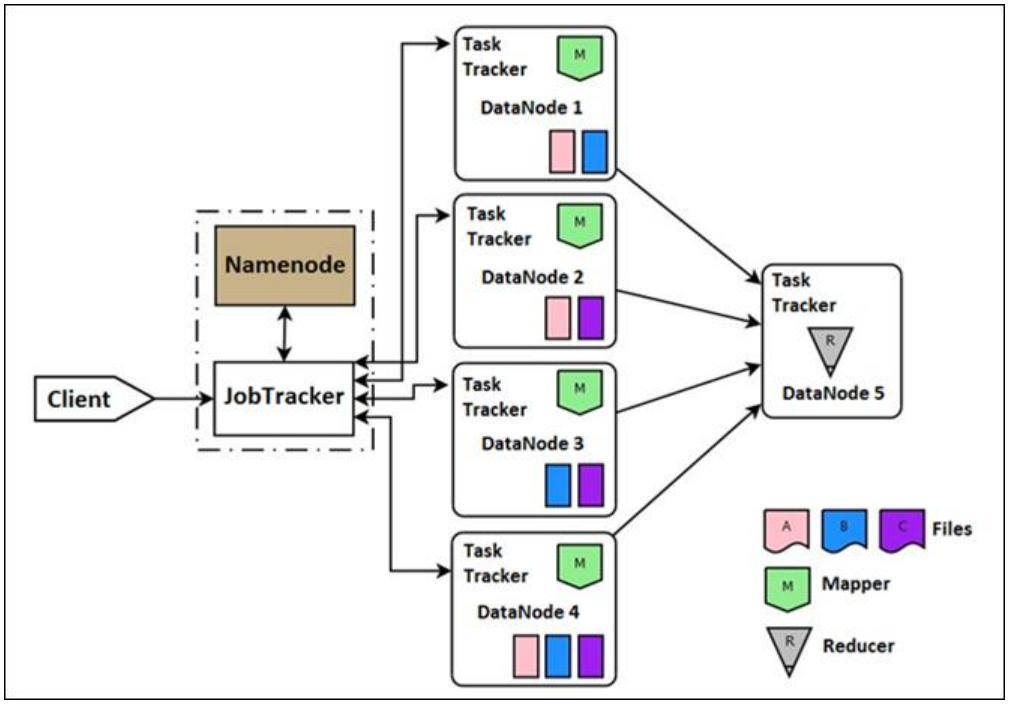
UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

hadoop
HDFS

.inf
UFRGS

Related Concepts





UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

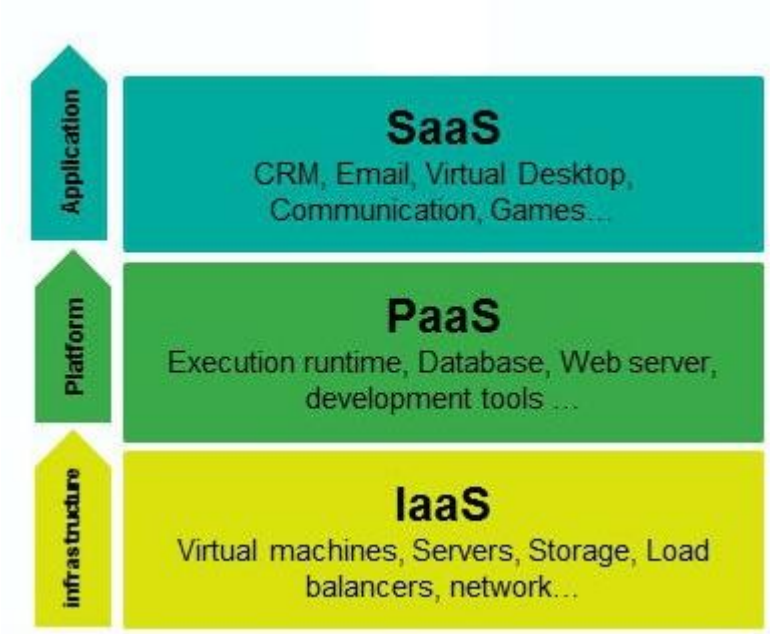
YAHOO!

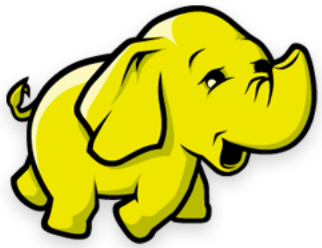
inf UFRGS

Microsoft Azure

Azure

HDINSIGHT





UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Results

- IaaS Cluster:
  - 1/2/3/4 Nodes.
  - D3 Type of node (4 Xeon E5-2660 | 14GB | 200GB).
  - Hadoop 1.2.1 .
  - Java 7.
  - Ubuntu 15.04 LTS.
  - Cost of 1 node: U$$0.376/h.

- PaaS Cluster:
  - 1/2/3/4 Nodes.
  - D3 Type of node (4 Xeon E5-2660 | 14GB | 200GB).
  - HD Insights 2.2 (Hadoop 2.6).
  - Java 7.
  - Ubuntu 12.04 LTS.
  - Cost of 1 node: U$$0.684/h.

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

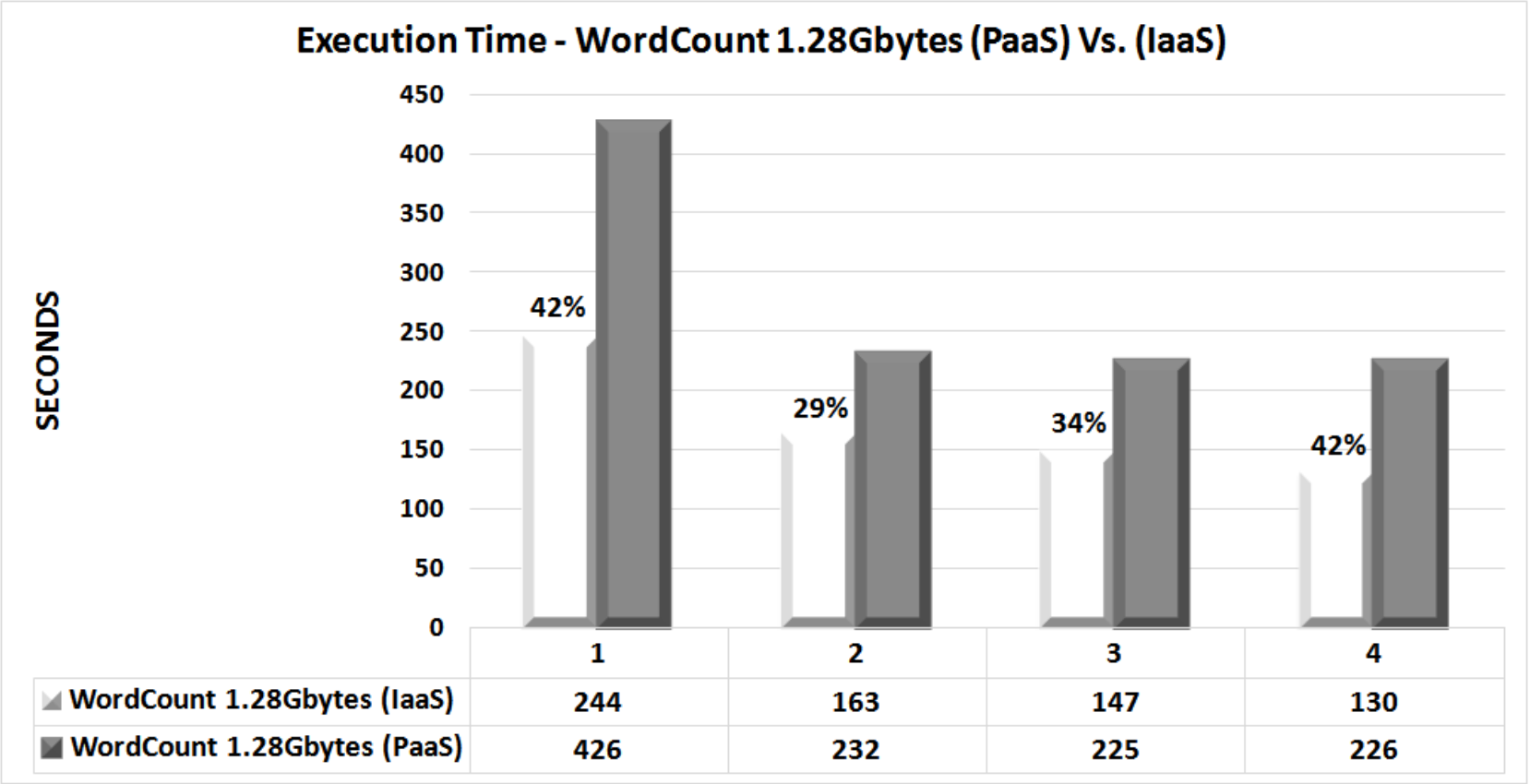
Results

- Benchmarks:
  - WordCount:
    - 1.28GB.
    - Generated by Linux random of 1M of numbers.
    - Need to read and write data on HDFS.
  - Pi:
    - 50 Maps generating data.
    - Need only to use the CPU.
- Metrics evaluated:
  - Time of execution.
  - SpeedUp.
  - Relation of cost and time of execution.

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

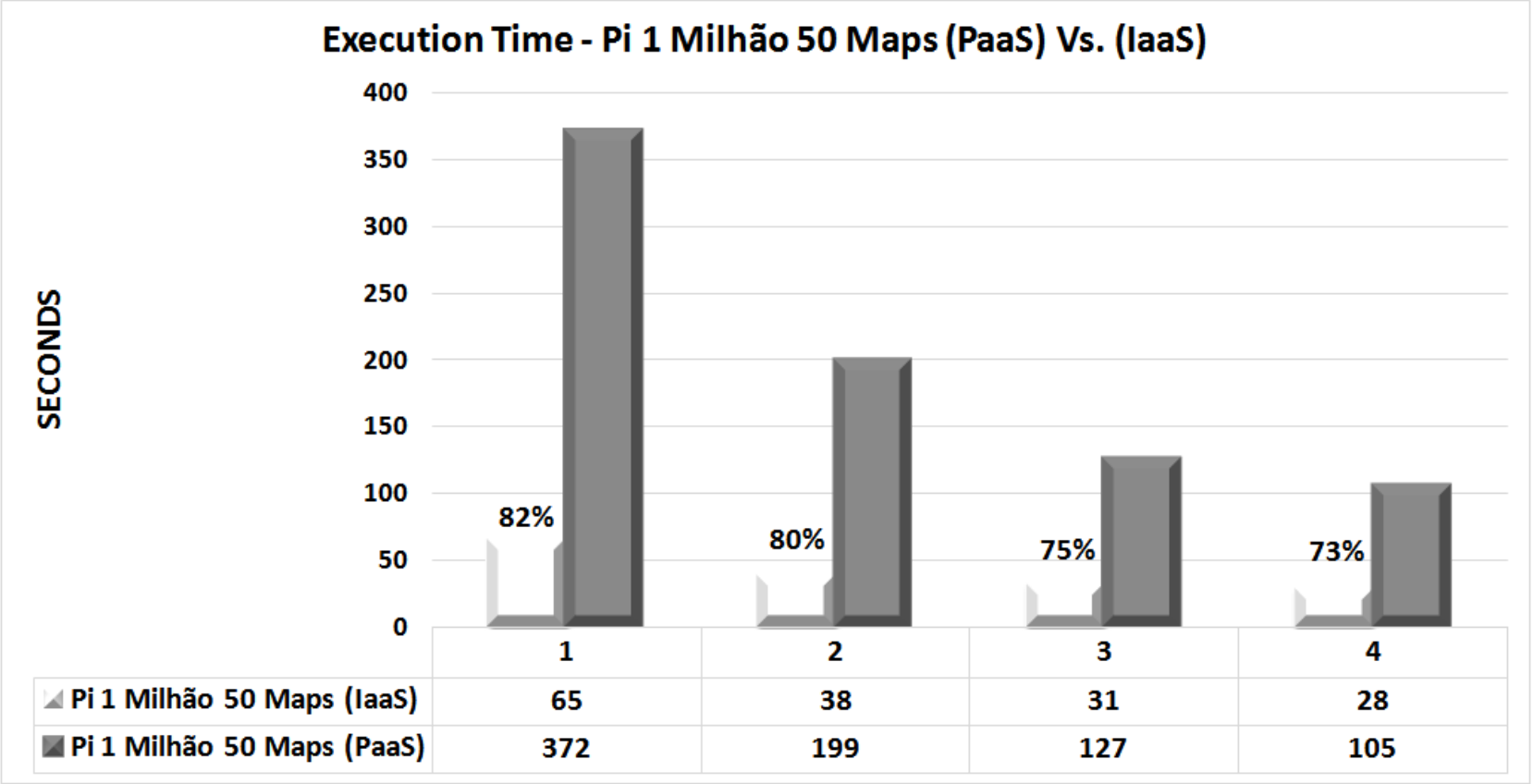
Results



UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

.inf UFRGS

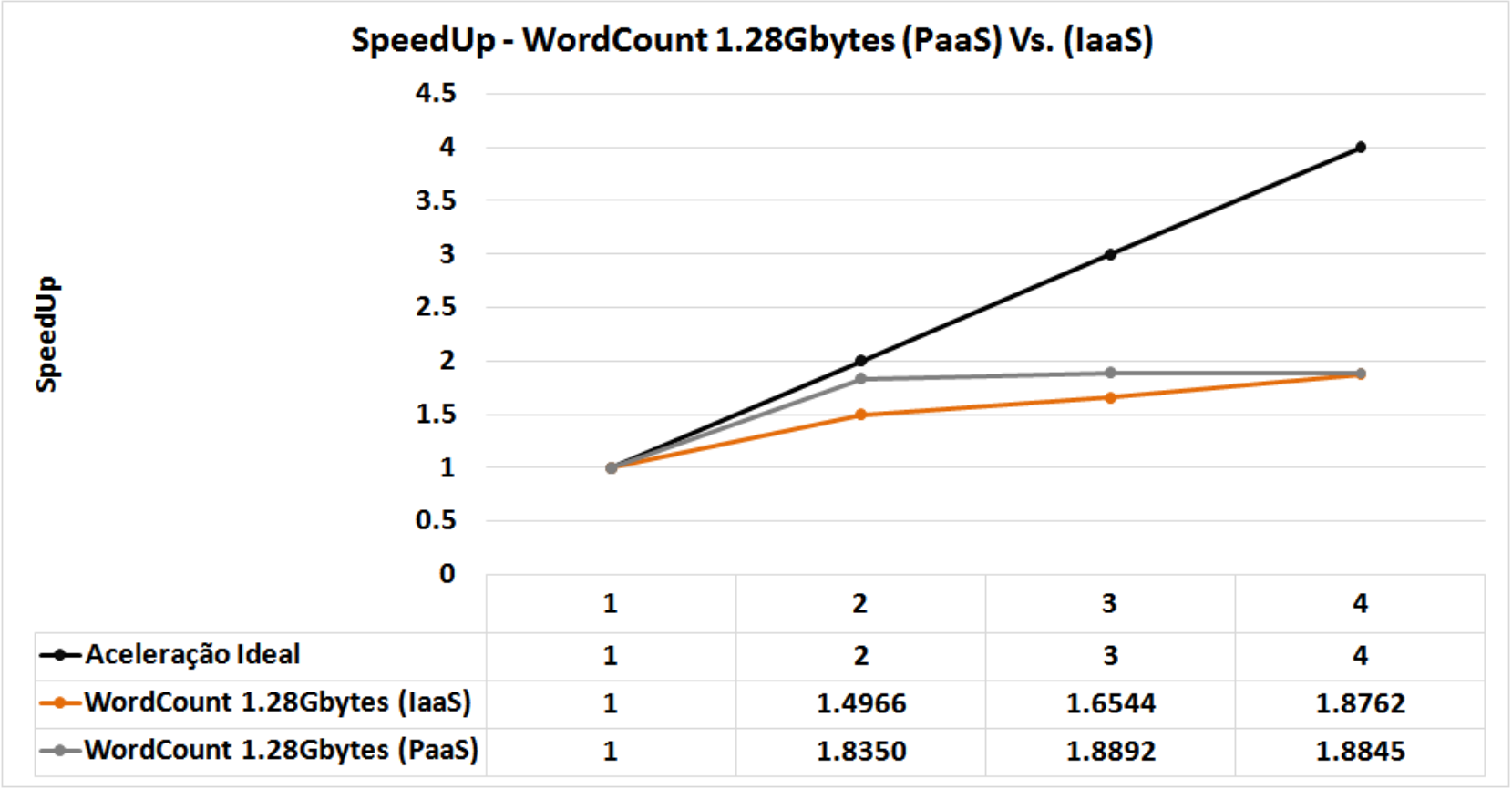
Results



UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Results

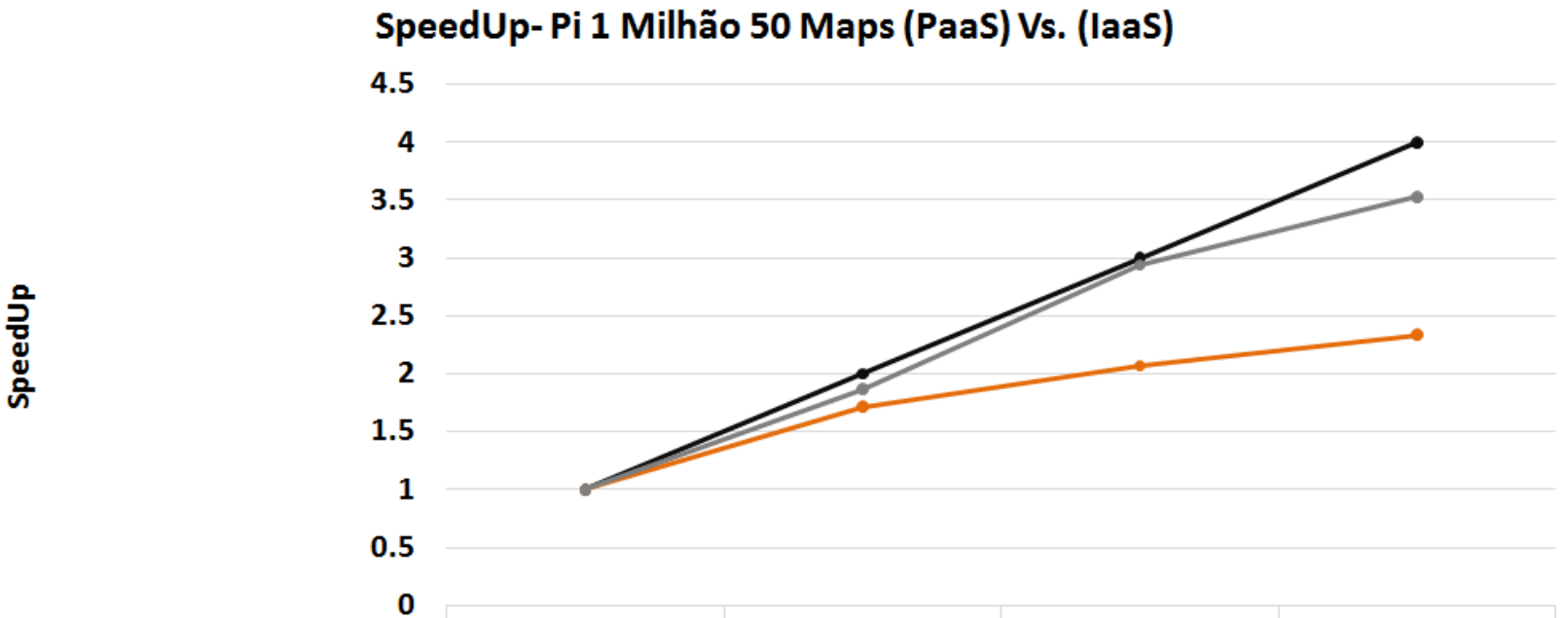


	1	2	3	4
Aceleração Ideal	1	2	3	4
WordCount 1.28Gbytes (IaaS)	1	1.4966	1.6544	1.8762
WordCount 1.28Gbytes (PaaS)	1	1.8350	1.8892	1.8845

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Results

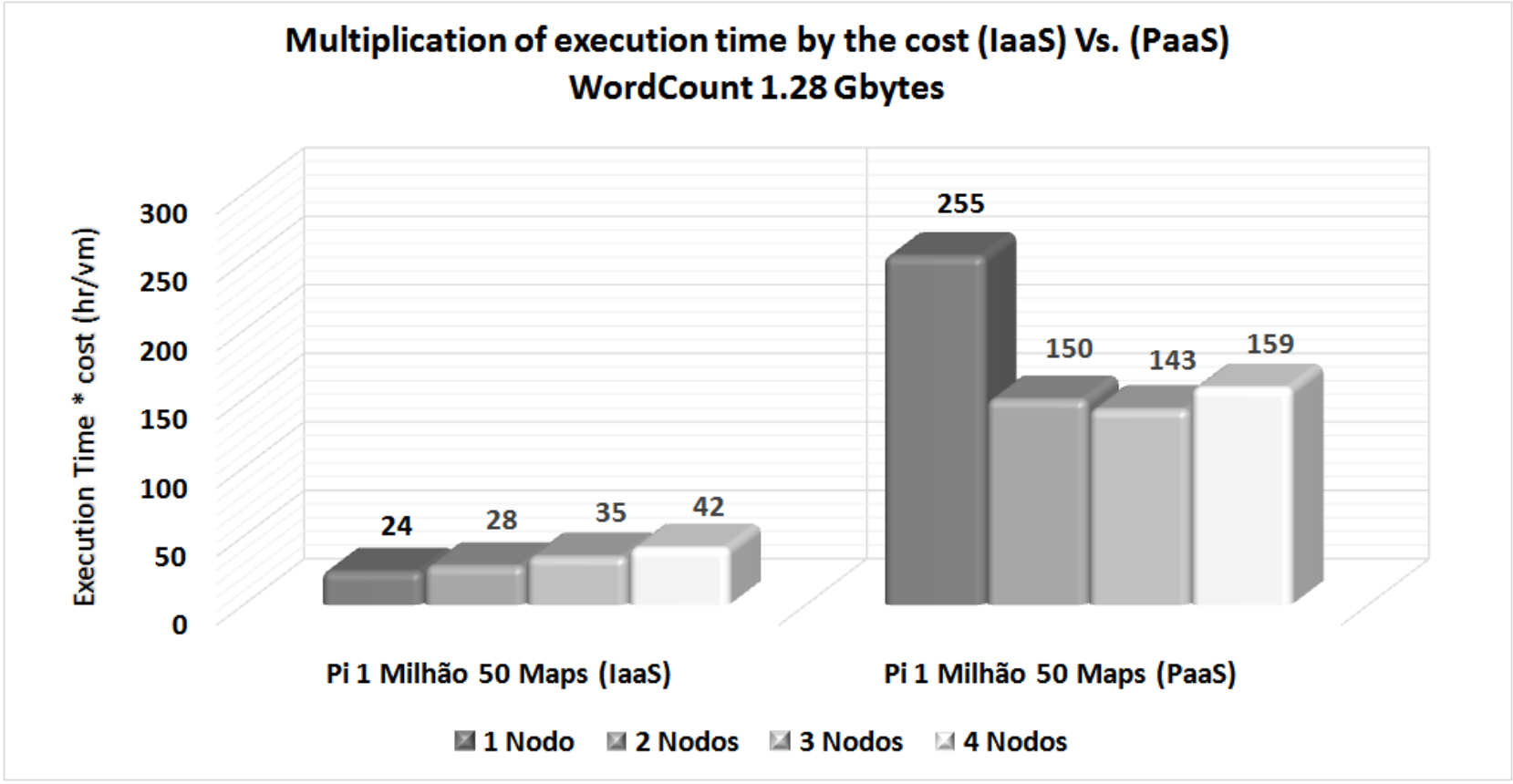


	1	2	3	4
Aceleração Ideal	1	2	3	4
Pi 1 Milhão 50 Maps (IaaS)	1	1.71513	2.07014	2.33451
Pi 1 Milhão 50 Maps (PaaS)	1	1.86533	2.93987	3.52905

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

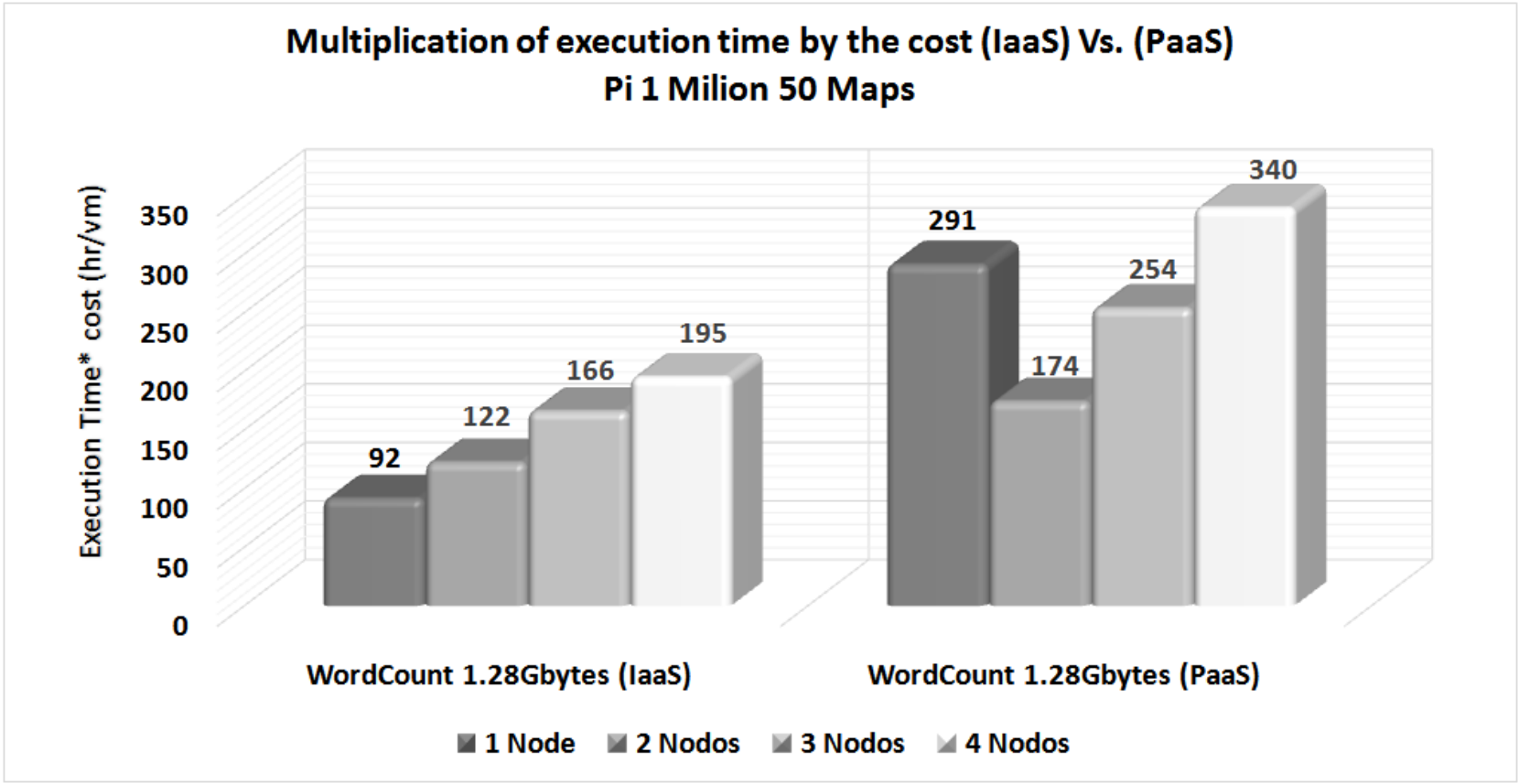
Results



UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Results



UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Conclusions

- IaaS present the best performance.
- PaaS present the best acceleration.
- In PaaS the service provider will make an effort to provide physical resources with high accuracy and low variability.
- This effort needs to be paid by the user. It reduces the efficiency of service PaaS.
- Use of Azure blob can be the reason of worse performance in PaaS service.

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

Future Works

- Use of Hadoop Yarn (2.x) in IaaS.
- Use of disk servisse Azure blobs in IaaS.
  - Compared to the use of local cache (SSD).
- Use of another types of VMs:
  - Type D (Data Intensive).
  - Type A (General Purposes).

References

- [1] Hp (2015) "*A BIG brother for your BIG data environment*". Acessado em Abril de 2015. Disponível em: http://h30499.www3.hp.com/t5/Business-Service-Management-BAC/A-BIG-brother-for-your-BIG-data-environment/ba-p/6284087.

- [2] White, T. (2012). Hadoop - The Definitive Guide, volume 1. OReilly Media, Inc., 3rd.

- [3] Peter M. Mell and Timothy Grance. 2011. SP 800-145. The NIST Definition of Cloud Computing. Technical Report. NIST, Gaithersburg, MD, United States.

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

.inf UFRGS

References

- [4] Chauhan, A; Fontama, Woody, B. “Introducing Microsoft Azure HDInsight”. Microsoft 2015.

- [5] Microsoft. “How Azure pricing works”. Disponível em: http://azure.microsoft.com/en-us/pricing/details/virtual-machines/. Acessado em : Agosto de 2015.

- [6] Schemmer, R.B., Anjos, J.C.S., Tibola, A.L., Barros, J.F., Geyer, C.F.R.; Framework Hadoop em Plataformas de Cloud e Cluster Computing. Escola Regional de Alto Desempenho (ERAD). pp.71,88. Abril 2015.

Questions ?

geyer@inf.ufrgs.br

raffael.schemmer@inf.ufrgs.br

INF/GPPD – UFRGS