

Estudo experimental do compressor BZIP2 em arquiteturas paralelas e distribuídas



Junior F. Barros

Mst. Raffael B. Schemmer

Prof. Mst. Julio C. S. Anjos

Prof. Dr. Claudio F. R. Geyer

- Introdução (Objetivos).
- Metodologia de avaliação.
- Resultados.
- Conclusões.
- Trabalhos Futuros.



- Compressão:
 - Redução no tamanho dos arquivos.
 - Com perdas - (JPEG/MPEG).
 - Sem perdas - (PNG/RAR).
- (Motivação) Técnica é capaz de:
 - Reduzir o consumo de rede e disco.
 - Executar em ambientes paralelos e distribuídos.
- Objetivo: Avaliar a escalabilidade do compressor BZIP2 em cenários multicore, cluster e cloud.



Metodologia de avaliação.

- BZIP2 obteve melhor relação de eficiência:
 - Comparado a outros compressores.
- Versão paralela (Multicore) e distribuída do BZIP2:
 - Não suportada por Linux debian e derivados.
- BZIP2 [Gilchrist 2012] é baseado no algoritmo Burrows-Wheeler [Burrows 1994].
 - Possui implementações paralelas e distribuídas.

- Três tipos de máquinas utilizadas:
 - multicore, cluster e cloud
- Benchmark utilizado:
 - *Silesia Compression Corpus*. [Mahoney 2014]
 - 12 formatos de arquivos.
- Testes realizados:
 - 3 execuções para cada formato de arquivo.
 - Execução em todas as máquinas.

- Máquina Multicore:
 - Core i5 3470s 2.9 GHz Ivy Bridge Quad-Core.
 - 4Gbytes de RAM.
 - 1TB HDD.
 - Linux CentOS 6 Kernel 2.6.



- Versão do BZIP2 (POSIX Threads).
- Código foi variado em 1,2,3 e 4 threads.

- Avaliação realizada:

- Aceleração de compressão e descompressão.



- Máquina Cluster (GPPD GradeP):
 - 10 Instâncias Pentium IV 2.8 GHz Single-Core
 - 2Gbytes de RAM cada
 - 1TB HDD cada
 - Linux Rocks 6 Kernel 2.6.
- Versão do BZIP2 (MPI).
- Código foi variado em 1 a 10 computadores.
- Avaliação realizada:
 - Aceleração de compressão e descompressão.



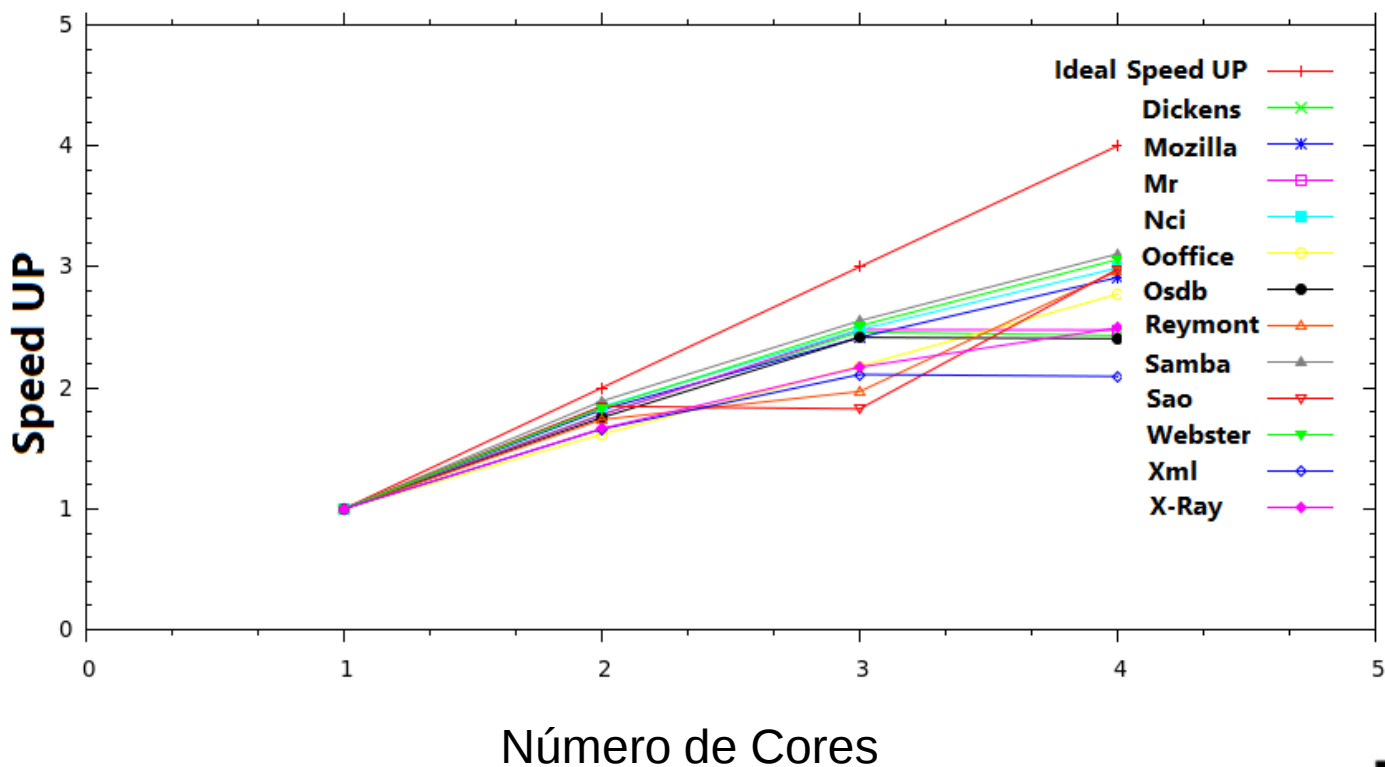


- Cenário Cloud (Microsoft Azure):
 - 10 Instâncias VMs A1.
 - (1 core) Xeon E5 2660 Sandy Bridge 2.2 GHz.
 - 1.75Gbytes de RAM cada
 - 30GB HDD cada
 - Linux Ubuntu 14.04 LTS

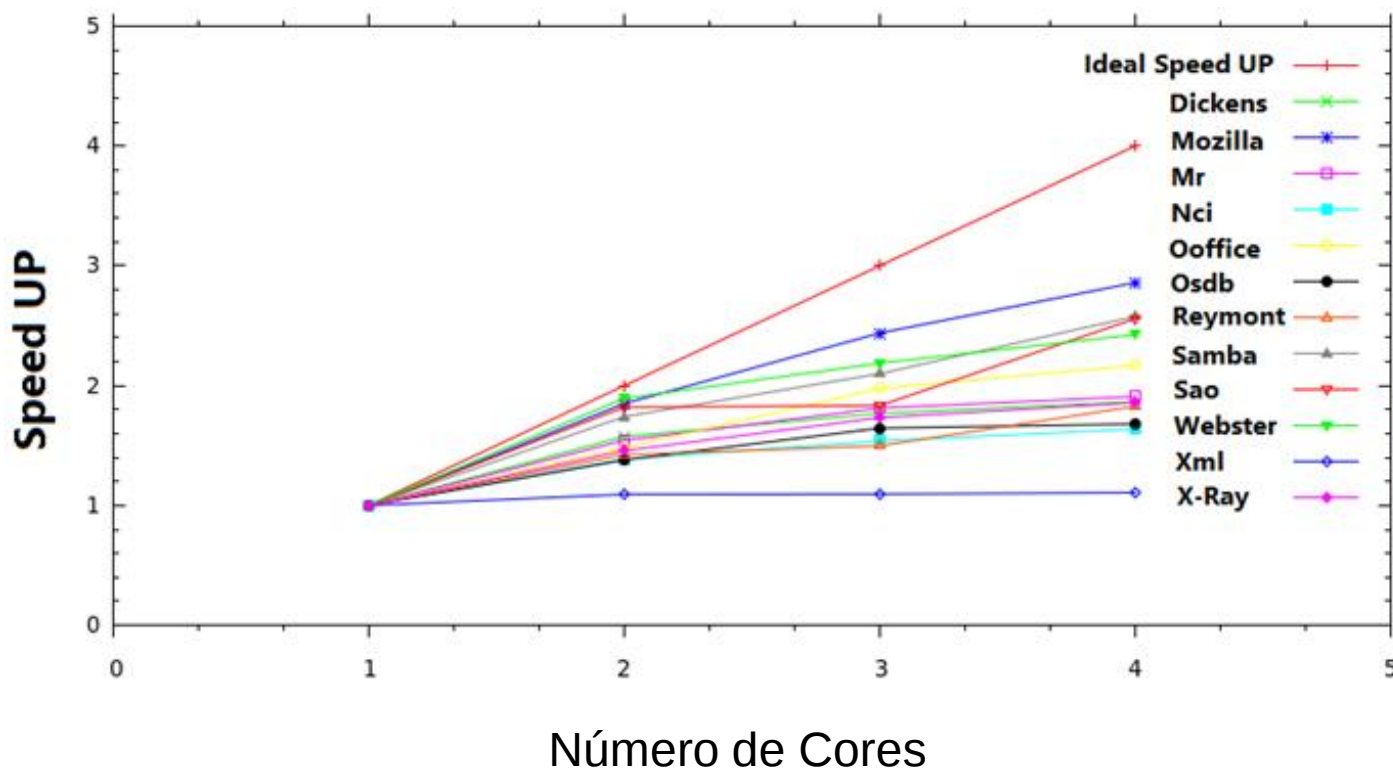


- Versão do BZIP2 (MPI).
- Código foi variado em 1 a 10 computadores.

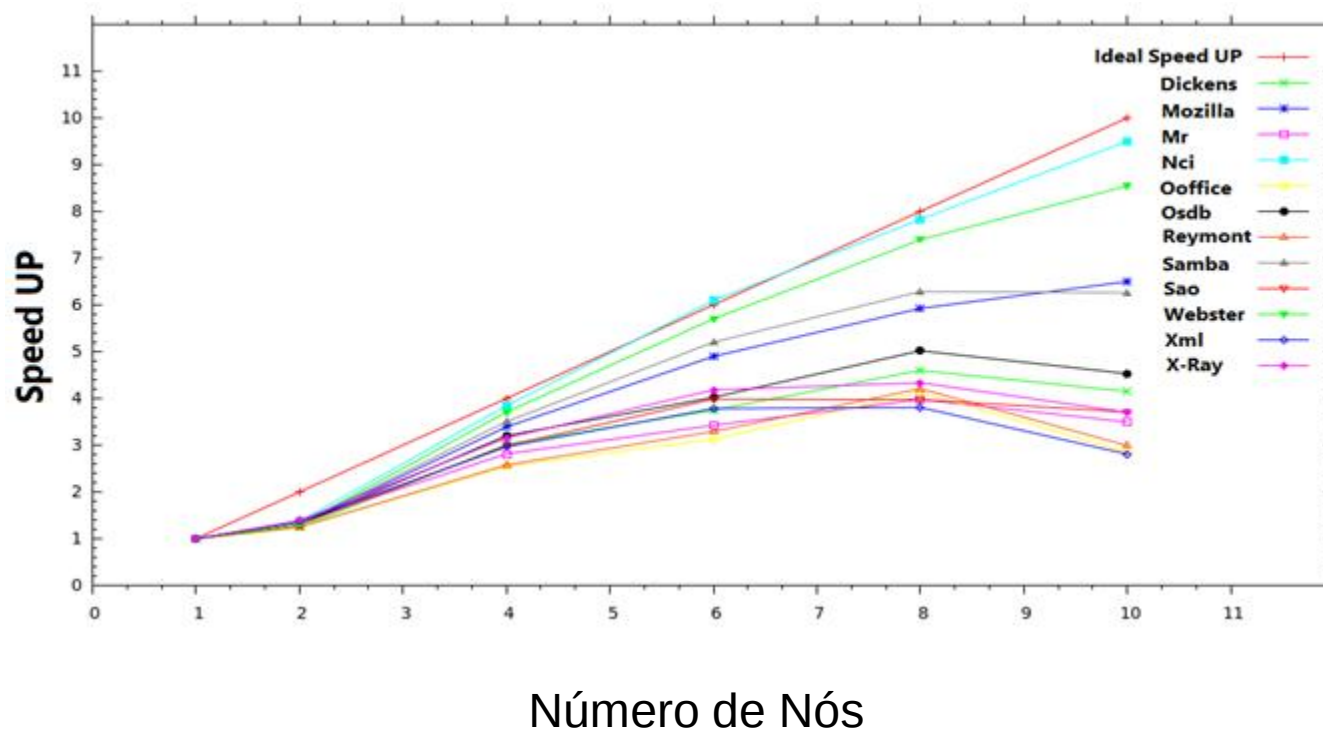
■ Cenário Multicore (compressão)



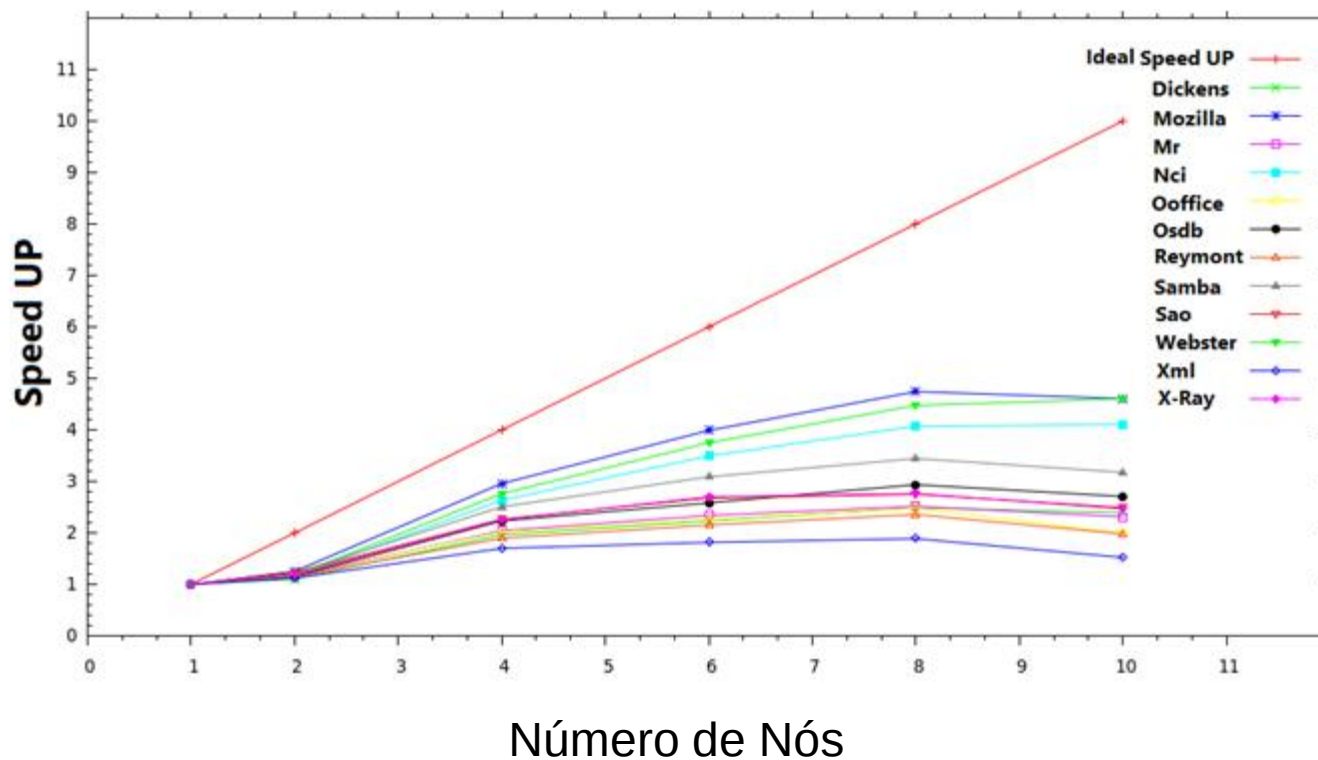
■ Cenário Multicore (descompressão)



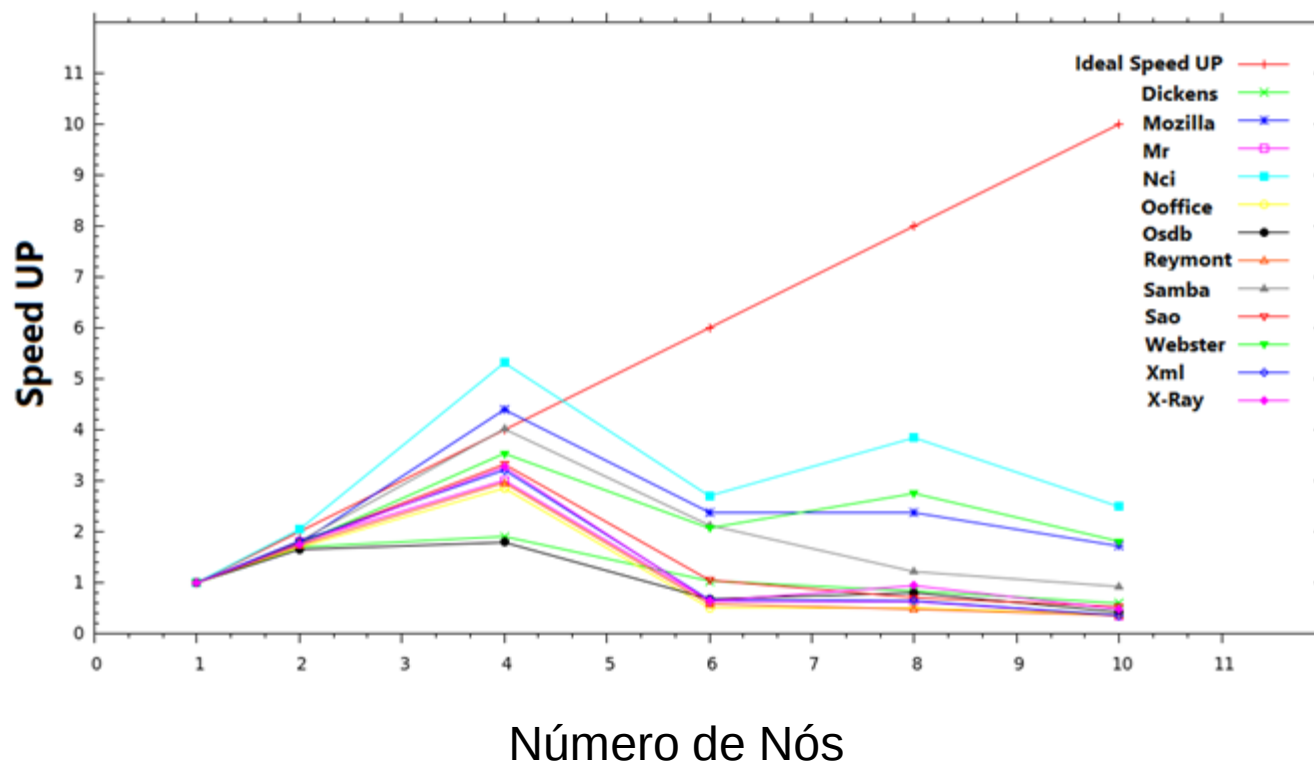
■ Cenário Cluster (compressão)



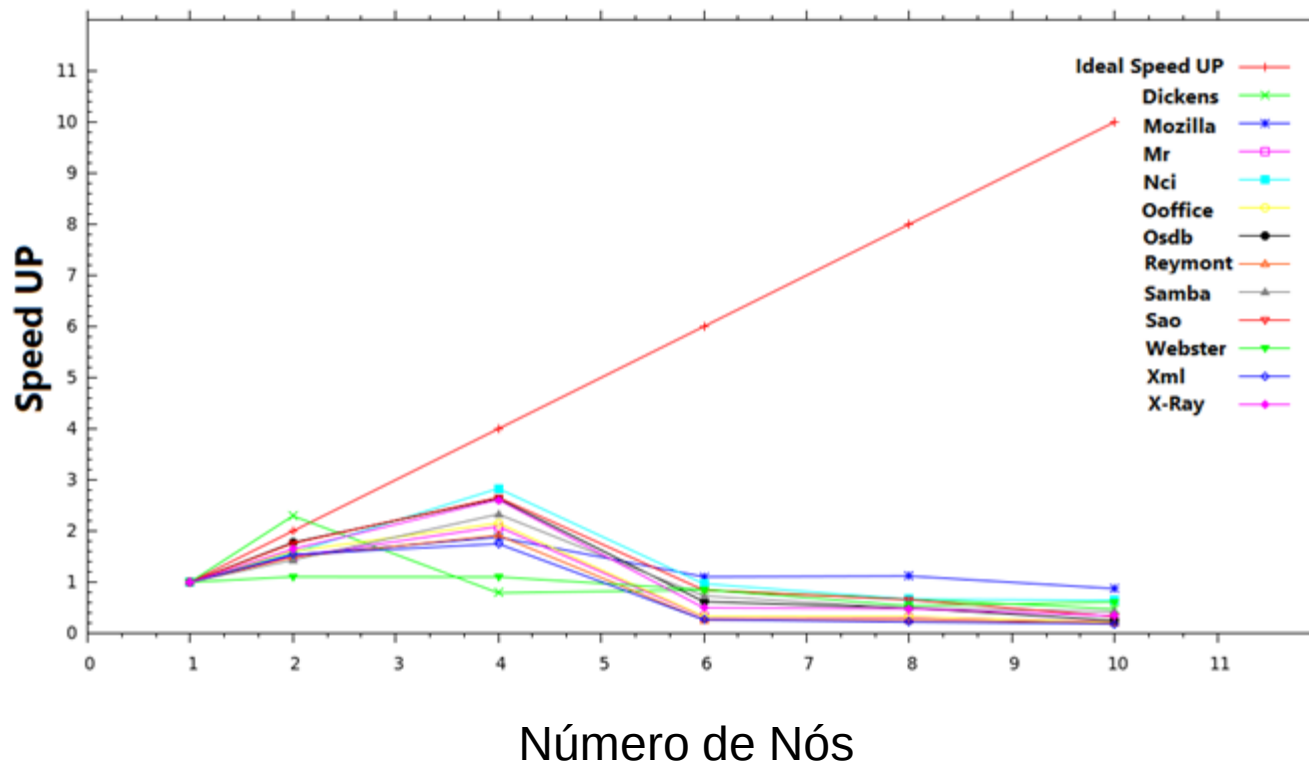
■ Cenário Cluster (descompressão)



■ Cenário Cloud (compressão)



■ Cenário Cloud (descompressão)



- BZIP2 se mostrou eficiente (compressão) e escalável.
- Cluster apresentou maior escalabilidade que cloud.
- Cenário não controlado da cloud pode:
 - Ser o causador da não escalabilidade

- Cloud pode receber e comprimir os dados:
 - Ganho de espaço.
 - Feito de forma paralela.
- Cliente pode receber dado comprimido:
 - Redução no largura de banda da rede.
 - BZIP2 possui excelente aceleração em Multicore.
- Técnica de compressão pode melhorar a performance e eficiência no uso de recursos de aplicações SaaS.

- Avaliar atrasos na rede da cloud (Azure).
 - Possíveis causadores da baixa escalabilidade.
- Utilizar afinidade entre nós na cloud (Azure):
 - Tentativa de aumento da aceleração.
- Uso de grandes arquivos de entrada.



Referências

- Burrows, M. (1994). “A block-sorting lossless data compression algorithm”. <http://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-124.pdf>.
- Gilchrist, J. (2007). “Parallel MPI BZIP2 (MPIBZIP2)”. <http://compression.ca/mpibzip2/>.
- Mahoney, M. (2014). “Silesia open source compression benchmark”. <http://mattmahoney.net/dc/silesia.html>.

Obrigado pela atenção!

Perguntas?



INF/UFRGS - GPPD
rafael.schemmer@inf.ufrgs.br