Universidade do Estado de Santa Catarina – Bacharelado em Ciência da Computação
Sistemas Operacionais – Prof. Rafael R. Obelheiro – Semestre: 2020.2

Trabalho 3 – Sistemas de Arquivos

Sobre o trabalho

- O trabalho é individual.
- Exercícios de implementação devem ser realizados na linguagem C, e serão testados no sistema operacional Linux. As submissões estarão sujeitas a controle de plágio usando ferramentas de detecção de similaridade.
- Exercícios teóricos devem ser submetidos em formato PDF. Para os exercícios de implementação deve ser submetido o código fonte, em formato compilável (ou seja, os arquivos `*.c`, e não listagens em PDF). Submeta um único arquivo ZIP contendo todas as suas respostas (teóricas e de implementação).
- O trabalho deverá ser entregue até **SEGUNDA-FEIRA, 22 DE MARÇO**. O Moodle aceitará submissões até 23h59min, sendo automaticamente bloqueado após a data limite. É **RESPONSABILIDADE DOS ALUNOS** garantir que o trabalho seja entregue no prazo.
- Em caso de dificuldades ou dúvidas na interpretação do trabalho, entre em contato com o professor (`rafael.obelheiro@udesc.br`).

Exercícios

1. O sistema de arquivos ext4 é provavelmente o mais usado hoje no Linux. O ext4 foi introduzido em 2008 e é o mais recente em uma linhagem que começou com o *Extended file system* (ext), introduzido em 1992, e que teve como sucessores o ext2 (introduzido em 1993) e o ext3 (introduzido em 2001). Até o ext3, todos os sistemas de arquivos da família usavam a estrutura tradicional de alocação indexada, com indireção simples, dupla e tripla. Uma das principais diferenças do ext4 é o uso de ***extents*** no lugar da alocação indexada. Outros sistemas de arquivos modernos que são amplamente usados, como NTFS, JFS e XFS, também fazem uso de *extents*.

   Faça uma pesquisa bibliográfica e responda às seguintes perguntas:

   (a) Como funciona a alocação baseada em *extents*, e como ela se diferencia da alocação indexada?
   (b) Qual a justificativa para introduzir um esquema de alocação diferente na evolução do ext3 para o ext4?
   (c) Até o ext3, o alocador de blocos retornava um bloco a cada vez que era invocado. Qual o problema que isso acarreta para o uso de *extents*? Que modificações são necessárias na alocação de blocos para que o esquema de *extents* funcione bem?
   (d) Qual dos esquemas introduz um *overhead* menor no acesso a arquivos pequenos (com 10 blocos ou menos), a alocação indexada ou a baseada em *extents*? Justifique sua resposta, considerando o caso em que os blocos são contíguos e o caso em que os blocos estão totalmente espalhados (i.e., não há blocos adjacentes).

   IMPORTANTE: você pode usar qualquer fonte como referência, mas **todas** as fontes consultadas devem ser listadas após as respostas.

2. Uma das decisões de projeto em sistemas de arquivos é a escolha do tamanho dos blocos lógicos. Como discutido em aula, blocos maiores oferecem maior eficiência na transferência de dados de/para o dispositivo de armazenamento, mas geram maior desperdício de espaço devido a fragmentação interna. Outra questão diz respeito à necessidade de escolher estruturas de metadados que comportem arquivos de diferentes tamanhos. Neste exercício você irá coletar dados do seu sistema de arquivos para raciocinar sobre essas questões.

   Primeiro, gere uma lista de arquivos em `/usr/bin` e `/usr/share` e seus respectivos tamanhos, usando o comando abaixo:

   ```
   $ find /usr/bin /usr/share -type f -exec ls -l {} \; 2>/dev/null | \
     awk '{print "\""$9"\"", $5}' | tee /tmp/lista-arqs
   ```

   Ao final da execução, o arquivo `/tmp/lista-arqs` conterá duas colunas, a primeira com os nomes de arquivos e a segunda com seus tamanhos (em bytes).

Com base nos dados coletados, responda às seguintes perguntas:

(a) Quantos arquivos foram encontrados?

(b) Qual o maior tamanho de arquivo observado? A que arquivo corresponde?

(c) Quantos arquivos com tamanho zero existem? A que porcentagem do total eles correspondem?

(d) Qual o tamanho médio de arquivo? Qual a porcentagem de arquivos com tamanho igual ou menor do que a média?

(e) Qual a mediana do tamanho de arquivo?

(f) Qual o menor tamanho de bloco necessário para que pelo menos 50% dos arquivos ocupem apenas um bloco? Qual a porcentagem de arquivos que ocuparia um bloco se esse tamanho fosse adotado? Quantos blocos (com o tamanho encontrado) ocuparia o maior arquivo?
NOTA: o tamanho de bloco deve ser uma potência de 2 (512 bytes, 1 KB, 2 KB, 4 KB, 8 KB, ...)

(g) Sabendo que um i-node no Linux possui 12 ponteiros diretos e indireção simples, dupla e tripla, e supondo que os ponteiros de disco (endereços de bloco) sejam de 32 bits, determine:

- a porcentagem de arquivos que não precisam de indireção (i.e., que usam apenas os ponteiros diretos);
- a porcentagem de arquivos que precisam de indireção simples;
- a porcentagem de arquivos que precisam de indireção dupla;
- a porcentagem de arquivos que precisam de indireção tripla.

(h) Considere que o sistema de arquivos aloca um número integral de blocos para cada arquivo. Por exemplo, se um arquivo tem 1337 bytes e o tamanho de bloco é de 1 KB, serão alocados dois blocos para esse arquivo, o que significa que ele ocupará 2 KB no disco, dos quais $2048 - 1337 = 711$ bytes serão desperdiçados por fragmentação interna. Sabendo disso, calcule o espaço desperdiçado por fragmentação interna (total e porcentagem) considerando os arquivos analisados.
DICA: você precisará saber o tamanho de bloco usado pelo seu sistema de arquivos. Para isso, use o comando `stat arq` (onde `arq` é um arquivo contido em `/tmp/lista-arqs`) e veja o valor de `IO Block`.

IMPORTANTE: você precisa incluir, no arquivo ZIP submetido no Moodle, o arquivo de dados (`/tmp/lista-arqs`) e qualquer planilha ou script usado para chegar às respostas.

3. Considere um sistema de arquivos que usa um mapa de bits para a gerência do espaço livre. Determine o tamanho (em bytes ou KB) ocupado para esse mapa com base nos parâmetros definidos na tabela abaixo.

Nome	Tamanho do bloco (KB)	Tamanho da área de dados (GB)
Alisson Felipe dos Santos	1	1
Alysson Rodrigo de Oliveira	2	2
Bruno Luis Vieira	4	3
Eduarda Cristina Rosa	8	4
Endrew Rafael Treptow Hang	16	5
Guilherme Araújo Lira de Menezes	1	6
Joao Marcelo Specki Xavier	2	7
José Eduardo Brandão	4	1
Josué Celeste do Nascimento	8	2
Leandro Andrei da Cunha	16	3
Marcelo Schena	1	4
Matheus Santana Carvalho	2	5
Murilo Francisco de Freitas	4	6
Nicolas Demantova Ribeiro	8	7
Pedro Kenzo Kawasaki Alves	16	1
Thiago Pimenta Barros Silva	1	2
Vinicius Gasparini	2	3