UNIVERSIDADE FEDERAL DE SANTA CATARINA

INE5412 - Sistemas Operacionais I Sistemas de Arquivos SimpleFS

Professores Giovani Gracioli e Márcio Castro 2023/02

1 Objetivos

Ao realizar este projeto, você irá:

- Aprender sobre as estruturas de dados e implementação de um sistema de arquivos similar ao do Unix.
- Aprender sobre recuperação de sistema de arquivos implementando uma varredura de bitmap de blocos livres.
- Desenvolver sua experiência em programação C++, utilizando classes e objetos extensivamente.
- Adquirir experiência ao desenvolver e empregar um regime rigoroso de testes.

2 Descrição

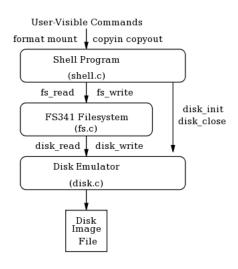
Neste projeto, você irá construir um sistema de arquivos. Este sistema de arquivos se chama **SimpleFS** e é muito similar à camada de inodo (*inode*) dos sistemas baseados em Unix. Você deverá escrever o código do sistema de arquivos considerando a descrição detalhada de projeto abaixo.

Para permitir que você desenvolva e teste um novo sistema de arquivos sem a preocupação de danificar discos e sistemas de arquivo existentes, você irá construir o SimpleFS usando um **emulador** de disco. Um emulador de disco tem todas as capacidades de um disco real: ele carrega e armazena dados persistentes em blocos. Ele até reclama e dá *crash* como um disco real. A única diferença é que não há discos reais girando: o emulador apenas armazena o dado do disco como um arquivo em um sistema de arquivos existente.

Para testar a sua implementação do SimpleFS, é fornecido um shell que permite você criar, formatar, ler e escrever arquivos no seu sistema de arquivos SimpleFS. Também é fornecido três exemplos de imagens de sistema de arquivos para que você não precise começar do zero.

2.1 Descrição do SimpleFS

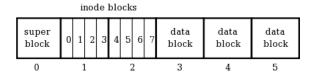
O SimpleFS tem 3 componentes principais: o shell, o sistema de arquivos em si e o disco emulado. Seu trabalho é implementar o componente intermediário: o sistema de arquivos. A Figura a seguir mostra como os componentes se relacionam entre si:



No nível mais alto o usuário fornece comandos digitados para um shell, instruindo-o a formatar ou montar o disco, e copiar dados para dentro ou fora do sistema do arquivo. O shell converte estes comandos digitados em operações de alto nível no sistema de arquivos, tais como fs_format, fs_mount, fs_read e fs_write. O sistema de arquivos é responsável por aceitar estas operações nos arquivos e convertê-los em operações de leitura e escrita de bloco simples no disco emulado, chamado de read e write. O disco emulado, por sua vez, armazena todos os seus dados em um arquivo de imagem no sistema de arquivos.

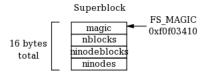
2.2 Projeto do sistema de arquivos SimpleFS

O SimpleFS tem o layout no disco a seguir. Ele assume que os blocos de disco têm um tamanho comum de 4 KB. O primeiro bloco do disco é um "superbloco" que descreve o layout do resto do sistema de arquivos. Um certo número de blocos que segue o "superbloco" contém as estruturas de dados inodos. Tipicamente, dez por cento do número total de blocos de disco são usados como blocos de inodos. Os blocos remanescentes no sistema de arquivos são usados como blocos de dados comuns, e ocasionalmente como blocos ponteiros indiretos. A Figura abaixo representa uma imagem do SimpleFS muito pequena:



Vamos examinar cada um destes tipos de blocos em detalhe.

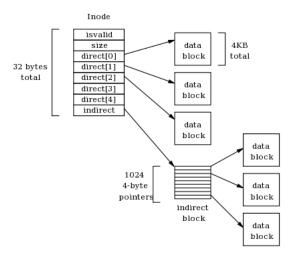
O superbloco descreve o layout do resto do sistema de arquivos:



Cada campo do superbloco é um inteiro de 4 bytes (32 bits). O primeiro campo é sempre o número "mágico" FS_MAGIC (0xf0f03410). A rotina de formatação coloca este número nos primeiros bytes do superbloco como um tipo de "assinatura" do sistema de arquivos. Quando o sistema de arquivos é montado, o SO procura por este número mágico. Se estiver correto, então assume-se que o disco contém um sistema de arquivos correto. Se algum outro número estiver presente, então a montagem falha, talvez porque o disco não esteja formatado ou contém algum outro tipo de dado.

Os campos remanescentes no superbloco descrevem o layout do sistema de arquivos. **nblocks** é o número total de blocos, que deveria ser o mesmo que o número de blocos no disco. **ninodeblocks** é o número de blocos reservados para armazenar inodos. **ninodes** é o número total de inodos nestes blocos. A rotina de formatação é responsável por escolher **ninodeblocks**: isto deve ser sempre 10 por cento de **nblocks**, arredondando pra cima. Note que a estrutura de dados do superbloco é pequena: apenas 16 bytes. O restante do bloco zero de disco é deixado sem ser usado.

Cada inodo se parece com o seguinte:



Cada campo do inodo é um campo de 4 bytes (32-bits). O campo **isvalid** é um se o inodo é válido (i.e se foi criado) e zero caso contrário. O campo **size** contém o tamanho lógico do dado do inode em bytes. Existem 5 ponteiros diretos para blocos de dados e um ponteiro para um bloco de dados indireto. Neste contexto, "ponteiro" significa simplesmente o número de um bloco onde dado pode ser encontrado. Um valor de zero pode ser usado para indicar um ponteiro de bloco nulo. Cada inodo ocupa 32 bytes. Então existem 128 inodos em cada bloco de inodo de 4 KB.

Note que um bloco de dados indireto é apenas um grande vetor de ponteiros para outros blocos de dados. Cada ponteiro é um inteiro de 4 bytes e cada bloco é 4 KB, então existem 1024 ponteiros por bloco. Os blocos de dados são simplesmente 4 KB de dados "crus".

Você provavelmente deve ter notado que algo está faltando: um bitmap de blocos livres! Um sistema de arquivos real manteria um bitmap de blocos livres no disco, gravando um bit para cada bloco que estivesse disponível ou em uso. Este bitmap seria consultado e atualizado toda vez que o sistema de arquivos precisasse adicionar ou remover um bloco de dados de um inodo.

SimpleFS não tem espaço em disco para um bitmap de blocos livres. Então, SimpleFS requer que o projetista do sistema operacional (i.e. VOCÊ) mantenha um bitmap de blocos livres em memória. Isto é, deve haver um vetor de inteiros, sendo um bit no vetor para cada bloco no disco, notando se o bloco está em uso ou disponível. Quando for necessário alocar um novo bloco para um arquivo, o sistema deve varrer através do vetor para localizar um bloco disponível. Quando um bloco for liberado, da mesma maneira, deve ser marcado no bitmap.

O que acontece quando memória é perdida? Suponha que o usuário faça algumas mudanças no sistema de arquivos SimpleFS, e então dê reboot no sistema. Sem um bitmap de blocos livres, o SimpleFS não consegue dizer quais blocos estão em uso e quais estão livres. Felizmente, esta informação pode ser recuperada lendo o disco. Cada vez que um sistema de arquivos SimpleFS é montado, o sistema deve construir um novo bitmap de blocos livres do zero varrendo por todos os inodos e gravando quais blocos estão em uso. (Isto é muito parecido com realizar um *fsck* toda vez que o sistema inicializa).

O SimpleFS se parece muito com a camada de inodo de sistemas operacionais baseados em Unix. Cada "arquivo" é identificado por um inteiro chamado de "inúmero". O inúmero é simplesmente um índice para dentro do vetor de estruturas inodo que começa no bloco um. Quando um arquivo é criado, o SimpleFS escolhe o primeiro inúmero disponível e o retorna para o usuário. Todas as referências subsequentes a este arquivo são feitas usando o inúmero. Usando o SimpleFS como uma base, você poderia facilmente adicionar outra camada de software que implementa nomes de arquivo e diretórios. No entanto, isto não fará parte do trabalho.

Agora que você sabe dos detalhes do sistema de arquivos a ser implementado, vamos discutir os detalhes técnicos do sistema de emulação.

2.3 Emulador de Disco

Foi fornecido a você um emulador de disco no qual armazenar seu sistema de arquivos. Este "disco" é na verdade armazenado como um grande arquivo no sistema de arquivos, assim você pode salvar dados em uma imagem de disco e então recuperá-los depois. Além disto, também foi fornecido a você três exemplos de imagens de disco com os quais você pode experimentar para testar seu sistema de arquivos.

Tal como um disco real, o emulador apenas permite operações em blocos inteiros de disco de 4 KB (DISK_BLOCK_SIZE). Você não pode ler ou escrever nenhuma unidade menor do que isto. O principal desafio de construir um sistema de arquivos é converter as operações requisitadas pelo usuário em quantidades arbitrárias de dados para operações em tamanhos fixos de blocos.

A interface do disco simulado é dada em disk.h, dentro da classe Disk:

```
class Disk {
public:
    static const unsigned short int DISK_BLOCK_SIZE = 4096;
    Disk(const char *filename, int nblocks);
    int size();
    void read(int blocknum, char *data);
    void write(int blocknum, const char *data);
    void close();
    ...
};
```

Antes de realizar qualquer operação no disco, você deve chamar o construtor e especificar uma imagem de disco (real) para armazenar os dados de disco e o número de blocos em um disco simulado. Se o construtor é

chamado em uma imagem de disco que já existe, o dado contido não será alterado. Quando você tiver terminado de usar o disco, chame **close** para liberar o arquivo. Estas duas chamadas já são feitas para você no shell, então você não deve mudá-las.

Assim que o disco for inicializado, você pode chamar size() para descobrir o numéro de blocos no disco. Como o nome sugere, read() e write() lê e escreve um bloco de dado no disco. Note que o primeiro argumento é um número de bloco. Por exemplo, a chamada obj.read(0,data) realizada a leitura dos primeiros 4 KB de dados no disco, assim como read(1,data) lê o próximo bloco de dados de 4 KB no disco. Toda vez que você invocar um read ou write, você deve se assegurar de que data aponte para 4 KB completos de memória.

Note que o disco tem algumas conveniências de programação que um disco real não tem. O disco real é um tanto "mimado" – se você enviar comandos inválidos, ele provavelmente irá dar crash no sistema ou se comportar de maneira estranha. Este disco simulado é mais "útil". Se você enviar um comando inválido, ele irá interromper o programa com uma mensagem de erro. Por exemplo, se você tentar ler ou escrever em um bloco de disco que não existe, você irá receber esta mensagem de erro:

```
ERROR: blocknum (592) is too big!
Abort (core dumped)
```

2.4 Interface do Sistema de Arquivos

Usando o disco simulado existente, você irá construir um sistema de arquivos funcional. Note que a interface para o sistema de arquivos e o código-base foram fornecidos. A interface é dada em fs.h pela classe INE5412_FS. Os métodos da interface são:

```
INE5412_FS(Disk *d);
int fs_format();
void fs_debug();
int fs_mount();

int fs_create();
int fs_delete( int inumber );

int fs_getsize( int inumber );

int fs_read( int inumber, char *data, int length, int offset );
int fs_write( int inumber, const char *data, int length, int offset );
```

Os métodos da interface devem funcionar da seguinte maneira:

• fs_debug – Varre um sistema de arquivos montado e reporta como os inodos e os blocos estão organizados. Se você conseguir escrever esta função, você já ganhou metade da batalha! Assim que você conseguir varrer e reportar as estruturas do sistema de arquivos, o resto é fácil. Sua saída do fs_debug deve ser similar ao seguinte:

```
superblock:

magic number is valid

1010 blocks on disk

101 blocks for inodes

12928 inodes total

inode 3:

size: 45 bytes

direct blocks: 103 194

inode 5:

size 81929 bytes

direct blocks: 105 109

indirect block: 210

indirect data blocks: 211 212 213 214 ...
```

- fs_format Cria um novo sistema de arquivos no disco, destruindo qualquer dado que estiver presente. Reserva dez por cento dos blocos para inodos, libera a tabela de inodos, e escreve o superbloco. Retorna um para sucesso, zero caso contrário. Note que formatar o sistema de arquivos não faz com que ele seja montado. Também, uma tentativa de formatar um disco que já foi montado não deve fazer nada e retornar falha.
- fs_mount Examina o disco para um sistema de arquivos. Se um está presente, lê o superbloco, constroi um bitmap de blocos livres, e prepara o sistema de arquivos para uso. Retorna um em caso de sucesso, zero caso contrário. Note que uma montagem bem-sucedida é um pré-requisito para as outras chamadas.
- fs_create Cria um novo inodo de comprimento zero. Em caso de sucesso, retorna o inúmero (positivo). Em caso de falha, retorna zero. (Note que isto implica que zero não pode ser um inúmero válido.)
- fs_delete Deleta o inodo indicado pelo inúmero. Libera todo o dado e blocos indiretos atribuídos a este inodo e os retorna ao mapa de blocos livres. Em caso de sucesso, retorna um. Em caso de falha, retorna 0.
- fs_getsize Retorna o tamanho lógico do inodo especificado, em bytes. Note que zero é um tamanho lógico válido para um inodo! Em caso de falha, retorna -1.
- fs_read Lê dado de um inodo válido. Copia "length" bytes do inodo para dentro do ponteiro "data", começando em "offset" no inodo. Retorna o número total de bytes lidos. O Número de bytes efetivamente lidos pode ser menos que o número de bytes requisitados, caso o fim do inodo seja alcançado. Se o inúmero dado for inválido, ou algum outro erro for encontrado, retorna 0.
- fs_write Escreve dado para um inodo válido. Copia "length" bytes do ponteiro "data" para o inodo começando em "offset" bytes. Aloca quaisquer blocos diretos e indiretos no processo. Retorna o número de bytes efetivamente escritos. O número de bytes efetivamente escritos pode ser menor que o número de bytes requisitados, caso o disco se torne cheio. Se o inúmero dado for inválido, ou qualquer outro erro for encontrado, retorna 0.

2.5 Interface Shell

Foi fornecido para você um shell simples que será usado para testar seu sistema de arquivos e o disco simulado. Na correção do trabalho, iremos usar o shell para testar o seu código, então certifique-se de testá-lo extensivamente. Para usar o shell, rode *simplefs* com o nome de uma imagem de disco, e o número de bloco desta imagem. Por exemplo, para usar o exemplo *image.5* dado abaixo, rode:

```
\% ./simplefs image.
5 5
```

Ou, para iniciar com uma imagem de disco nova, apenas dê um novo nome de arquivo e número de blocos:

```
\% ./simplefs mydisk 25
```

Assim que o shell iniciar, você pode usar o comando help para listar os comandos disponíveis:

```
simplefs> help
Commands are:

format

mount
debug

create
delete < inode >

cat < inode >

copyin < file >< inode >

copyout< inode >< file >

help

quit

exit
```

A maioria dos comandos correspondem de uma maneira muito próxima à interface de sistema de arquivos. Por exemplo, format, mount, debug, create e delete chamam os métodos correspondentes no sistema de arquivos.

Certifique-se de que você chame esses métodos em uma ordem sensata. Um sistema de arquivos deve ser formatado uma vez antes que possa ser usado. Igualmente, deve ser montado antes de ser escrito ou lido.

Os comandos complexos são cat, copyin, e copyout. cat lê um arquivo inteiro do sistema de arquivos e mostra no console, assim como o comando Unix de mesmo nome. copyin e copyout copiam um arquivo do sistema de arquivos Linux local para dentro do seu sistema de arquivos emulado. Por exemplo, para copiar o arquivo de dicionário para o inodo 10 no seu sistema de arquivos, faça o seguinte:

```
simplefs> copyin /usr/share/dict/words 10
```

Note que estes três comandos executam fazendo um grande número de chamadas para fs_read e fs_write para cada arquivo a ser copiado.

No arquivo *shell.cc* a classe File_Ops fornece uma implementação base para *copyin e copyout*, na qual as chamadas para as operações do sistema de arquivos já estão sendo feitas.

3 Começando

Baixe o <u>código-fonte</u> no Moodle e construa-o com make. No Moodle (dentro do arquivo compactado) ainda estão algumas imagens de disco de exemplo para você começar (image.5, image.20 e image.200). O nome de cada imagem de disco te informa quantos blocos existem em cada imagem. Cada imagem contém alguns documentos e arquivos familiares. Assim que você conseguir ler o que está nessas imagens, você deve prosseguir, escrevendo e modificando-os.

shell e disk são dados e completamente implementados, e fs.h e fs.cc são um "esqueleto" esperando pelo seu trabalho. Foram fornecidas as primeiras linhas de fs_debug para te dar a idéia inicial. Para começar, contrua e execute simplefs e chame debug para ler e ter como saída o superbloco:

```
% make
% ./simplefs image.5 5
simplefs> debug
superblock:
    magic number is valid
    5 blocks
    1 inode blocks
    128 inodes
```

4 Notas de Implementação

Sua tarefa é implementar o SimpleFS como descrito acima, preenchendo a implementação de fs.cc utilizando a linguagem C++. Você não precisa alterar quaisquer outros módulos de código. Dentro da classe $INE5412_FS$, foram definidas algumas estruturas de dados de amostra para você começar. Estas podem ser encontradas em fs.h. Para começar, nós definimos um número de constantes que você irá usar. A maioria destas devem ser auto-explicativas:

```
static const unsigned int FS_MAGIC = 0xf0f03410;
static const unsigned short int INODES_PER_BLOCK = 128;
static const unsigned short int POINTERS_PER_INODE = 5;
static const unsigned short int POINTERS_PER_BLOCK = 1024;
```

Note que *POINTS_PER_INODE* é o número de ponteiros diretos em cada estrutura de inodo, enquanto que POINTERS_PER_BLOCK é o número de ponteiros a ser encontrados em um bloco indireto.

O superbloco e estruturas inodo são facilmente traduzidas das Figuras acima e declaradas como *inner classes*:

```
class fs_superblock {
    unsigned int magic;
    int nblocks;
    int ninodeblocks;
    int ninodes;
};

class fs_inode {
    int isvalid;
    int size;
    int direct[POINTERS_PER_INODE];
    int indirect;
};
```

Atente-se que vários inodos podem caber em um bloco de disco. Um pedaço de 4 KB de memória correspondente a um bloco indireto pareceria com o seguinte:

```
int pointers[1024];
```

Finalmente, cada bloco de dados é apenas dados binários "crus" usados para armazenar o conteúdo parcial de um arquivo. Um bloco de dados pode ser especificado em C++ como simplesmente um vetor de 4096 bytes:

```
char data[4096];
```

Um bloco de disco "cru" pode ser usado para representar 4 tipos de dado: um superbloco, um bloco de 128 inodos, um ponteiro indireto de bloco, ou um simples bloco de dados. Isto implica em um problema de engenharia de software: como transformamos o dado "cru" retornado por *Disk::read* em cada um dos 4 tipos de blocos de dados?

C/C++ fornece um pedaço de sintaxe para exatamente este problema. Pode-se declarar uma union de cada um dos nossos 4 tipos de dados. Uma union se parece com uma struct, mas força todos os seus elementos a compartilhar o mesmo espaço de memória. Você pode pensar como uma union de vários tipos diferentes, todas sobrepostas em cima das outras.

Aqui está um exemplo. Cria-se uma $union\ fs_block$ que representa as quatro maneiras diferentes de interpretar dados de disco crus:

```
union fs_block {
    class fs_superblock super;
    class fs_inode inodes[128];
    int pointers[1024];
    char data[4096];
};
```

Note que o tamanho de uma união fs_block será exatamente 4 KB: o tamanho do maior membro da união. Para declarar uma variável do tipo: union fs_block :

```
union fs_block block;
```

Agora, pode-se usar *Disk::read* para carregar o dado cru do bloco zero. Passe ao *Disk::read* a variável block.data, que se assemelha a um array de caracteres:

```
disk.read(0, block.data);
```

Mas, podemos interpretar este dado como se fosse um objeto do tipo *classe superblock* acessando a parte superior da união. Por exemplo, para extrair o número mágico do super bloco, podemos fazer isto:

```
x = block.super.magic;
```

Por outro lado, suponha que queiramos carregar o bloco de disco 59, assumindo que é um bloco indireto e então examinar o quarto ponteiro. Novamente, usaríamos *Disk::read* para carregar o dado cru:

```
disk.read(59, block.data);
```

Mas então usaríamos a parte pointer da união da seguinte maneira:

```
x = block.pointer[4];
```

A união oferece uma maneira conveniente de visualizar o mesmo dado em múltiplas perspectivas. Quando nós carregamos dado do disco, é apenas um pedaço de dados cru de 4 KB (block.data). Mas, assim que carregado, a camada do sistema de arquivos sabe que este dado tem alguma estrutura. A camada do sistema de arquivos pode visualizar o mesmo dado sob outra perspectiva escolhendo outro campo na união (block.super por exemplo).

5 Conselhos Gerais

- 1. Implemente as funções aproximadamente em ordem. Nós intencionalmente apresentamos as funções do sistema de arquivos em ordem de dificuldade. Implemente debug, format, e mount primeiro. Assegure-se que você seja capaz de acessar as imagens de disco de exemplo fornecidos. Então, realize a criação e deleção de inodos sem se preocupar com dados de blocos. Implemente a leitura e teste novamente com as imagens do disco. Se todo o resto estiver funcionando, então tente fs_write.
- 2. Dividir para conquistar. Trabalhe duro para separar ações comuns em funções simples. Isto irá simplificar seu código dramaticamente. Por exemplo, você irá frequentemente precisar carregar e salvar estruturas de inodos individuais por número. Isto envolve um pouco de computação para transformar um inúmero em um número de bloco, e assim por diante. Então, faça dois métodos para fazer exatamente isto:

```
void inode_load( int inumber, class fs_inode *inode ) { ... } void inode_save( int inumber, class fs_inode *inode ) { ... }
```

Agora, sempre que você precisar carregar ou salvar uma estrutura de inodo, chame estes métodos. Sempre que você se encontrar escrevendo trechos de códigos muito similares repetidamente, separe-os em métodos menores.

- 3. Teste condições de contorno. Nós vamos certamente testar seu código analisando as exceções. Assegure-se de testar e consertar condições de contorno antes de entregar o trabalho. Por exemplo, o que acontece se fs_create descobre que uma tabela de inodo está cheia? Deve, de maneira clara, retornar um código de erro. Certamente não deve dar crash no programa ou mutilar o disco! Pense criticamente sobre outras possíveis condições de contorno tal como o fim de um arquivo ou um disco cheio.
- 4. Não se preocupe com desempenho. Você será avaliado pela exatidão, não desempenho. Aliás, ao longo deste trabalho, você descobrirá que um simples acesso de arquivo pode facilmente resultar em dezenas ou centenas de acessos individuais ao disco. Entenda porque isto ocorre, mas não se preocupe com a otimização.

6 Entregando o Trabalho

Este trabalho deve ser entregue pelo Moodle. Entregas após a data limite não serão aceitos!

Para entregar sua atividade, comprima todos os arquivos .cc e .h, juntamente com um Makefile em um único arquivo e submeta no moodle.

7 Avaliação

A avaliação se dará em 4 fases:

1. Avaliação de compilação: compilar o código enviado. Caso haja erros de compilação, a nota do trabalho será automaticamente zerada.

- 2. Avaliação de execução: para validar que a solução executa corretamente sem falhas de segmentação. Caso haja falhas de segmentação, a nota é zerada. Será também avaliado o uso de variáveis globais (-5 pontos) e vazamentos de memória (-20%).
- 3. Avaliação da organização do código: busca-se nesta fase avaliar a organização do código orientado a objetos e o seguimento das diretrizes do trabalho. Alguns itens para avaliação são: (i) funcionamento do programa; (ii) saída do programa (conforme especificação); (iii) clareza do código (utilização de comentários e nomes de variáveis adequadas); (iv) compilação sem warnings; (v) quantidade de funcionalidades do sistema de arquivos implementadas corretamente.
- 4. A quarta fase consiste na apresentação do trabalho em dia e horário agendado pelo professor. Durante as apresentações, o professor irá avaliar o conhecimento individual dos alunos sobre os conteúdos teóricos e práticos vistos em aula e sobre a solução adotada no tra'balho. A nota atribuída à cada aluno i no trabalho $(NotaTrabalho_i)$ será calculada da seguinte forma, onde A_i é a nota referente à apresentação do aluno i e S é a nota atribuída à solução do trabalho:

$$NotaTrabalho_i = \frac{A_i \times S}{10} \tag{1}$$

Plágio não será tolerado em nenhuma hipótese ao longo dos trabalhos, acarretando em nota 0 a todos os envolvidos.

8 Auxílios

8.1 Sáida esperada da função debug para o disco.20

fs-shell> debug
superblock:
20 blocks
3 inode blocks
384 inodes
inode 2:
size: 27160 bytes
direct blocks: 4 5 6 7 8
indirect block: 9
indirect data blocks: 13 14
inode 3:
size: 9546 bytes
direct blocks: 10 11 12