

# Google File System

Danilo Silva Marshall

Érika R. C. de Almeida

---

# Tópicos abordados

- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Tópicos abordados

- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Sistemas de arquivos

- O que é um sistema de arquivos?
  - É um conjunto de rotinas e estruturas lógicas que permitem ao sistema operacional gerenciar arquivos (como são estruturados, nomeados, acessados, usados, protegidos, implementados)
  - Fornece uma interface cliente através da qual a informação pode ser manipulada sem necessidade de conhecer detalhes de implementação

# Sistemas de arquivos - Linux

- EXT2
- EXT3
- ReiserFS
- XFS
- JFS

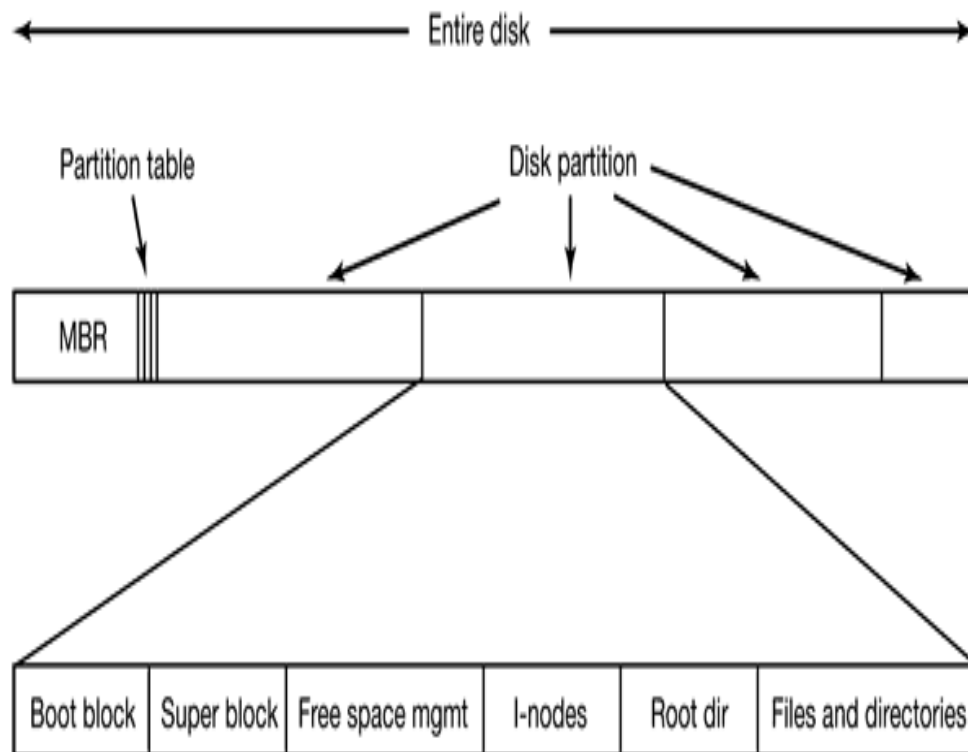
# Sistemas de arquivos - Windows

- FAT16 (MS-DOS)
- FAT32
- NTFS

# Sistemas de arquivos - Implementação

- Sistemas de arquivos são armazenados em disco
- Setor 0 do disco é chamado de MBR (Master Boot Record) e usado para fazer o boot
- O final do MBR contém a tabela de partições

# Sistemas de arquivos - Implementação



- Boot block: Carregar o sistema operacional da partição
- Superblock: Parâmetros do file system
- Free space: Bitmap ou listas ligadas
- I-nodes: Um por arquivo, contém a descrição
- Root dir: Diretório topo da árvore do sistema de arquivos
- Files e directories



# Tópicos abordados

- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Sistemas de arquivos distribuídos

- Fornece os mesmos serviços e recursos de um sistema de arquivos convencional com a diferença que o acesso à informação remota ou local pode ser realizado de forma transparente ao usuário

# Sistemas de arquivos distribuídos

- Vantagens:
  - Maior disponibilidade de informação
  - Visão única que todos os clientes têm do sistema de arquivos
- Desvantagem:
  - Complexidade

# Sistemas de arquivos distribuídos

- Requisitos:
  - Transparência
  - Segurança
  - Escalabilidade
  - Heterogeneidade
  - Disponibilidade
  - Confiabilidade

# Sistemas de arquivos distribuídos

- Exemplos:

- Andrew File System (AFS): o primeiro a oferecer alta escalabilidade. Desenvolvido na Universidade Carnegie-Mellon
- CODA: Desenvolvido a partir do AFS, mas com foco em alta disponibilidade
- Network File System (NFS): Sun Microsystems (1985)
- Global File System: Propriedade da Red Hat Inc, portado para Linux. Tem como objetivo promover alto desempenho para um número pequeno de clientes
- E ainda...

# Tópicos abordados

- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Google File System

- Sistema de arquivos desenvolvido e utilizado no Google
- Específico para o seu ambiente
- Capaz de suportar o enorme volume de dados processado diariamente
- Uso de muitas máquinas de baixo custo com capacidade de armazenamento alta e suporte a muitos acessos

# Google File System

- Elementos em comum com os outros sistemas:
  - Desempenho
  - Escalabilidade
  - Confiabilidade
  - Disponibilidade
  - Suporte a operações usuais em arquivos: create, delete, close, open, read, write

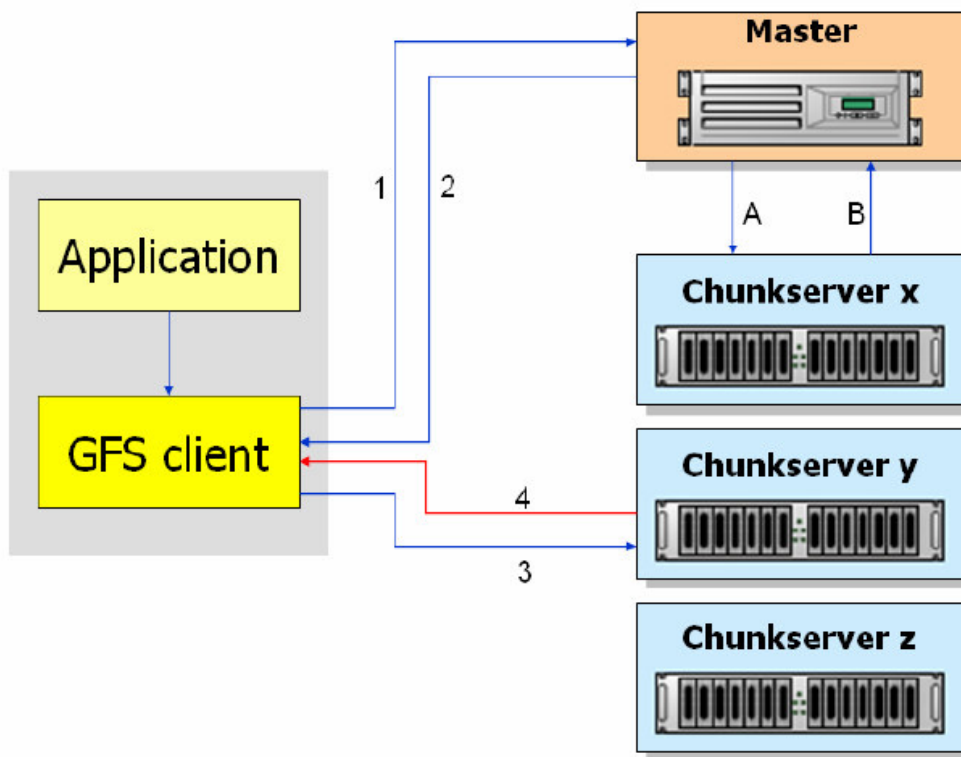


# Google File System

- Diferenciais:
  - Mecanismos de monitoramento constante para detecção de erros e recuperação automática
  - Maior quantidade de Append (escrita seqüencial no final do arquivo) ao invés de escrita aleatória (praticamente inexistente)
  - Operações de Snapshot e Record Append em arquivos

# Google File System - Arquitetura

- Um cluster do GFS consiste de um único *master* e múltiplos *chunkserver*s



As setas 1 e 2 indicam a comunicação entre o Cliente e o Mestre para a obtenção do metadado desejado a partir dos dados necessários.

As setas 3 e 4 indicam a comunicação entre o Cliente e o Chunkserver para a obtenção dos dados.

As setas A e B indicam a comunicação entre o Mestre e o Chunkserver para que o Mestre dê instruções ao Chunkserver e obtenha sua situação.

# Google File System - Arquitetura

- Arquivos:
  - Divididos em partes de tamanho fixo de 64 Mb, chamadas de *chunk*
  - Cada *chunk* é identificado por um *chunk handle* de 64 bits
  - Por confiabilidade, cada *chunk* é replicado em múltiplos *chunkservers* (por padrão, 3 réplicas)

# Google File System - Arquitetura

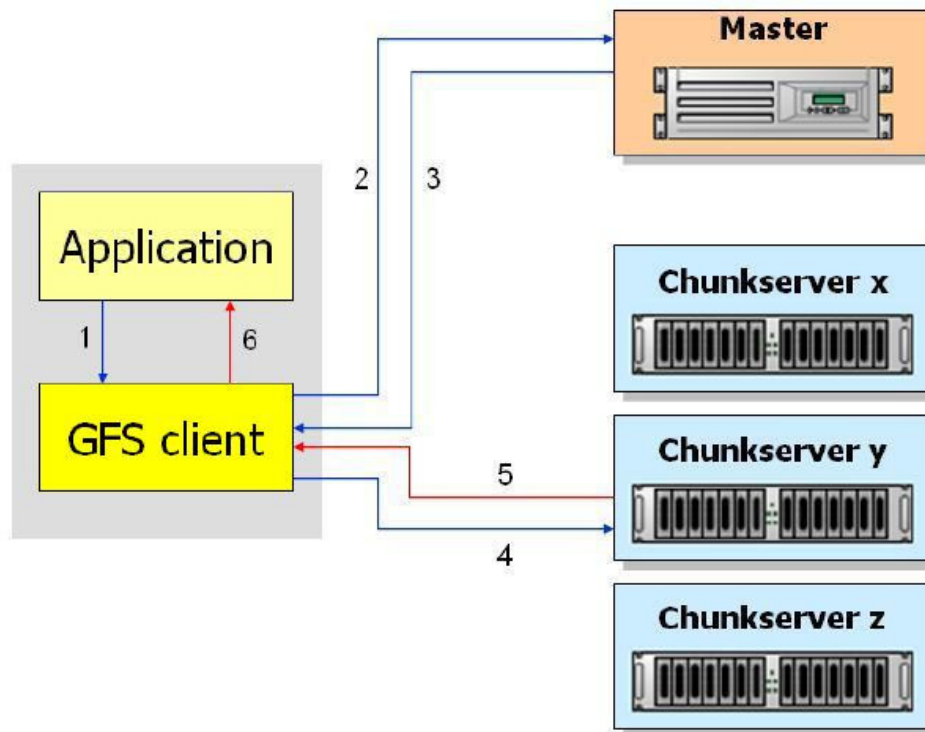
- *Master*:
  - Não é diretamente envolvido na transferência de arquivos
  - Responsável por armazenar *metadata* do sistema de arquivos. *Metadata* inclui: namespace, controle de acesso, mapeamento de arquivos para os *chunks* e a localização dos *chunks*
  - Controla o empréstimo de *chunks*, *garbage collection*, migração de *chunks* entre *chunkservers*
  - Comunicação periódica com os *chunkservers* para dar instruções e saber o estado

# Google File System - Arquitetura

- *Chunkservers:*
  - Armazena os chunks em seus discos como arquivos convencionais do Linux
  - Não faz cache de dados (assim como acontece com o cliente)

# Google File System - Operações

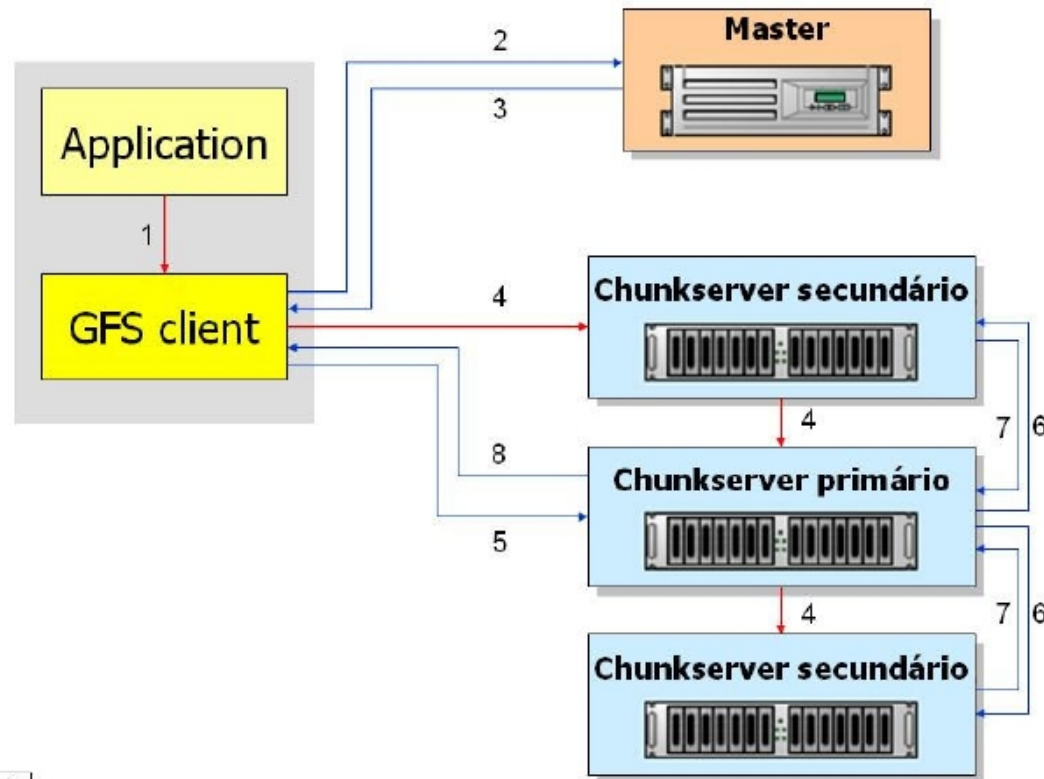
- Read



1. A aplicação requisita dados do sistema de arquivos ao Cliente
2. O cliente entra em contato com o Mestre para obter informações do metadado
3. O Mestre fornece o chunk em que se encontra a informação desejada e os Chunkservers que possuem réplicas do chunk (O cliente faz cache desta informação)
4. O cliente entra em contato com um dos Chunkservers que mantém réplica do chunk e envia o identificador do chunk
5. Transmissão do chunk (do Chunkserver para o cliente)
6. Leitura do chunk pela aplicação

# Google File System - Operações

- Write e Record Append:



# Google File System - Conclusão

- O Google File System demonstra as qualidades essenciais para suportar processamento de dados em larga escala



# Tópicos abordados

- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Gmail File System - Linux

- Projeto inicial do site <http://richard.jones.name/>
- Utiliza programa do tipo *fuse* que é usado para montar sistema de arquivos diversos
- Feito um programa em python
- Site de referência até hoje

# Gmail File System - Linux

- Vantagens:
  - Utilização de uma conta de gmail como um file system
  - O GmailFS suporta a maioria das operações, como ler, escrever, abrir, fechar, links simbólicos, links e renomear. Isso quer dizer que você pode seus comandos favoritos para operar os arquivos guardados no Gmail. Exemplo: cp, ls, ln, mv, rm, grep, etc

# Gmail File System – Linux

- Desvantagens:
  - Instalação complexa (mas não impossível!)
  - Não pode ser acessado de qualquer computador
  - Falhas de segurança
  - Gravação em blocos

# Gmail File System – Linux

- Instalação:
  - Necessários quatro arquivos:
    - Fuse
    - Fuse-python
    - Gmailfs
    - Libgmail

# Gmail File System – Linux (Instalação)

- FUSE

```
$ tar zxvf fuse-1.3.tar.gz  
$ cd fuse-1.3  
$ ./configure && make  
$ su  
# make install
```

Problema:

- Kernel source e Permissão

# Gmail File System – Linux (Instalação)

- Fuse-python

```
$ tar zxvf fuse-python.tar.gz  
$ cd fuse-python  
$ python setup.py build  
$ su  
# python setup.py install
```

Problema:

- Kernel e Permissão

# Gmail File System – Linux (Instalação)

- Gmailfs e libgmail

- Problemas de permissão apenas

GmailFS

```
$ tar zxvf gmailfs.tar.gz
$ cd gmailfs
$ su
# cp gmailfs /bin ; cp mount.gmailfs /sbin
```

libgmail

```
$ tar zxvf libgmail-0.0.8.tgz
$ cd libgmail-0.0.8
$ su
# cp libgmail.py constants.py
  /usr/lib/python2.3/site-packages/
```



# Gmail File System - Linux

- Após concluir a instalação, só irá faltar montar:

```
# mkdir /mnt/gmail  
# mount -t gmailfs /bin/gmailfs.py /mnt/gmail -o  
username=usuario,password=senha_do_email,fsname=Z  
dkfE
```

# Tópicos abordados

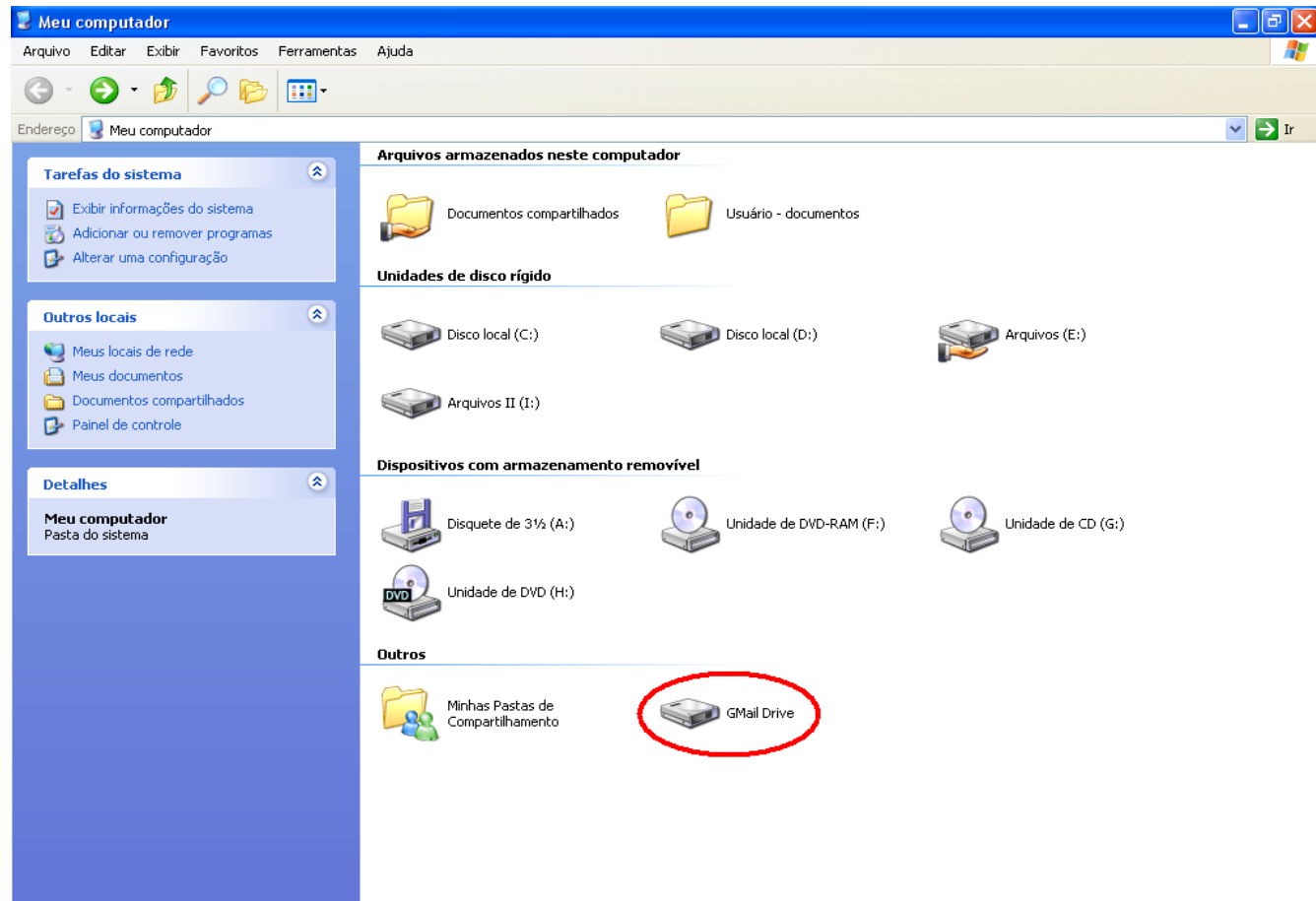
- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Gmail File System - Windows

- Mais conhecido como Gmail Drive
- Arquivo de instalação facilmente encontrado na Internet
- Funcionamento é mostrado a seguir

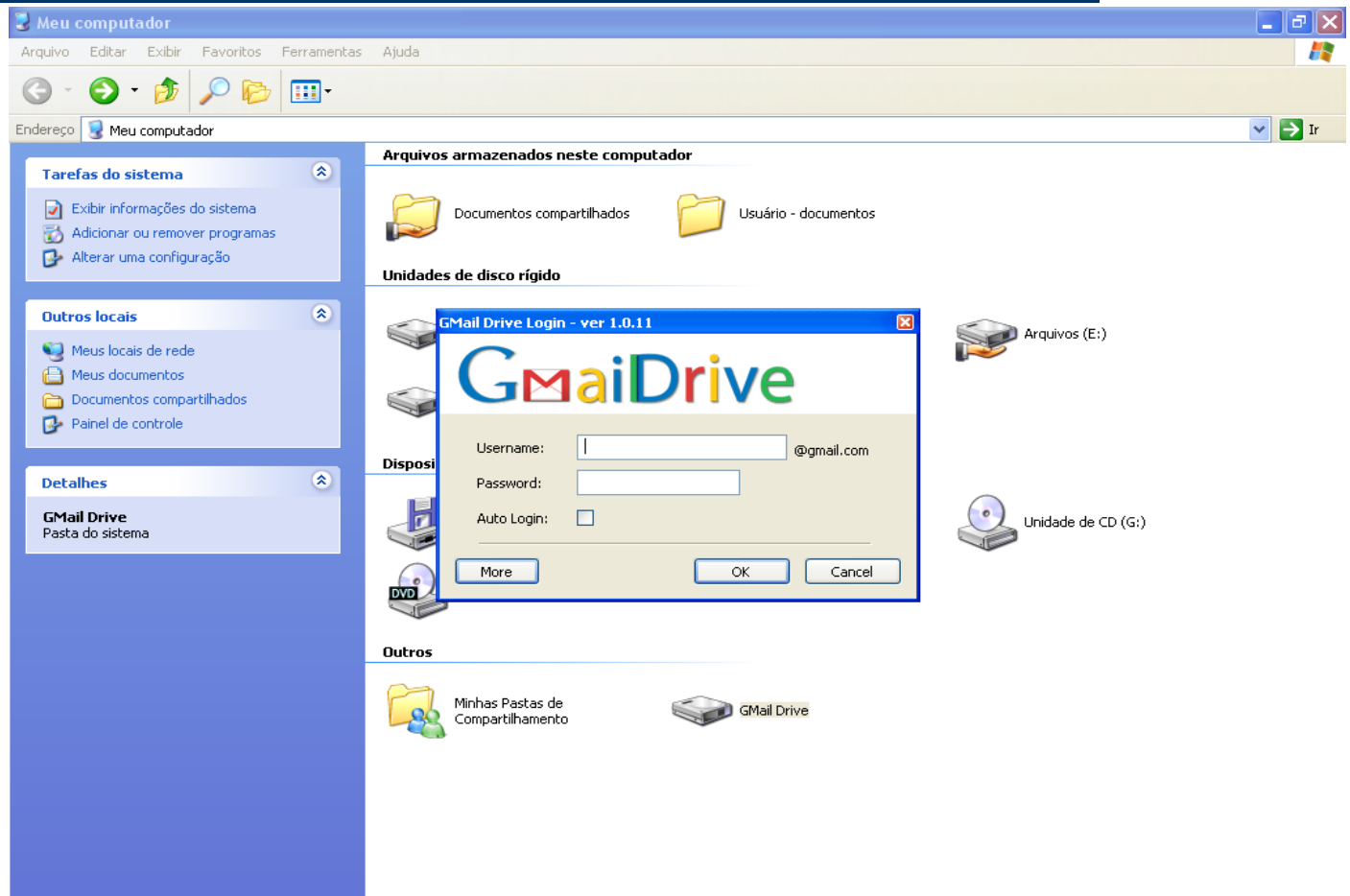
# Gmail File System - Windows

- Chamada:



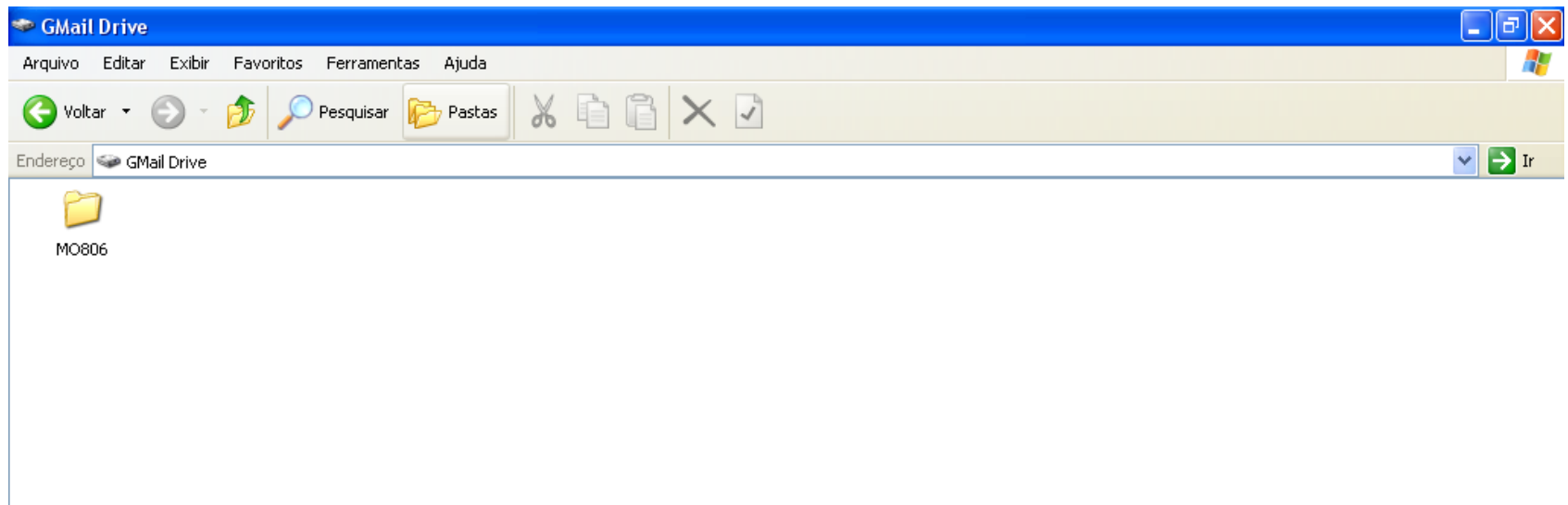
# Gmail File System - Windows

- Login:



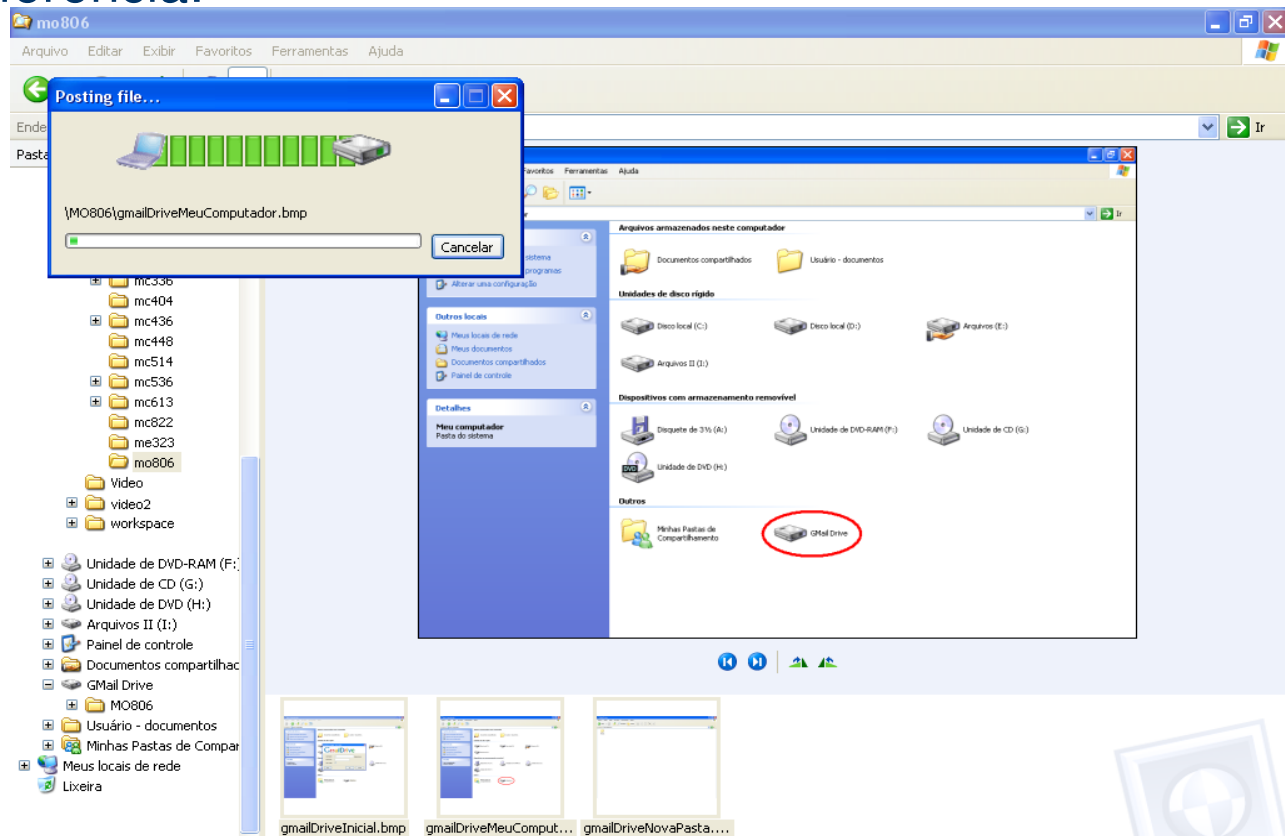
# Gmail File System - Windows

- Nova pasta:



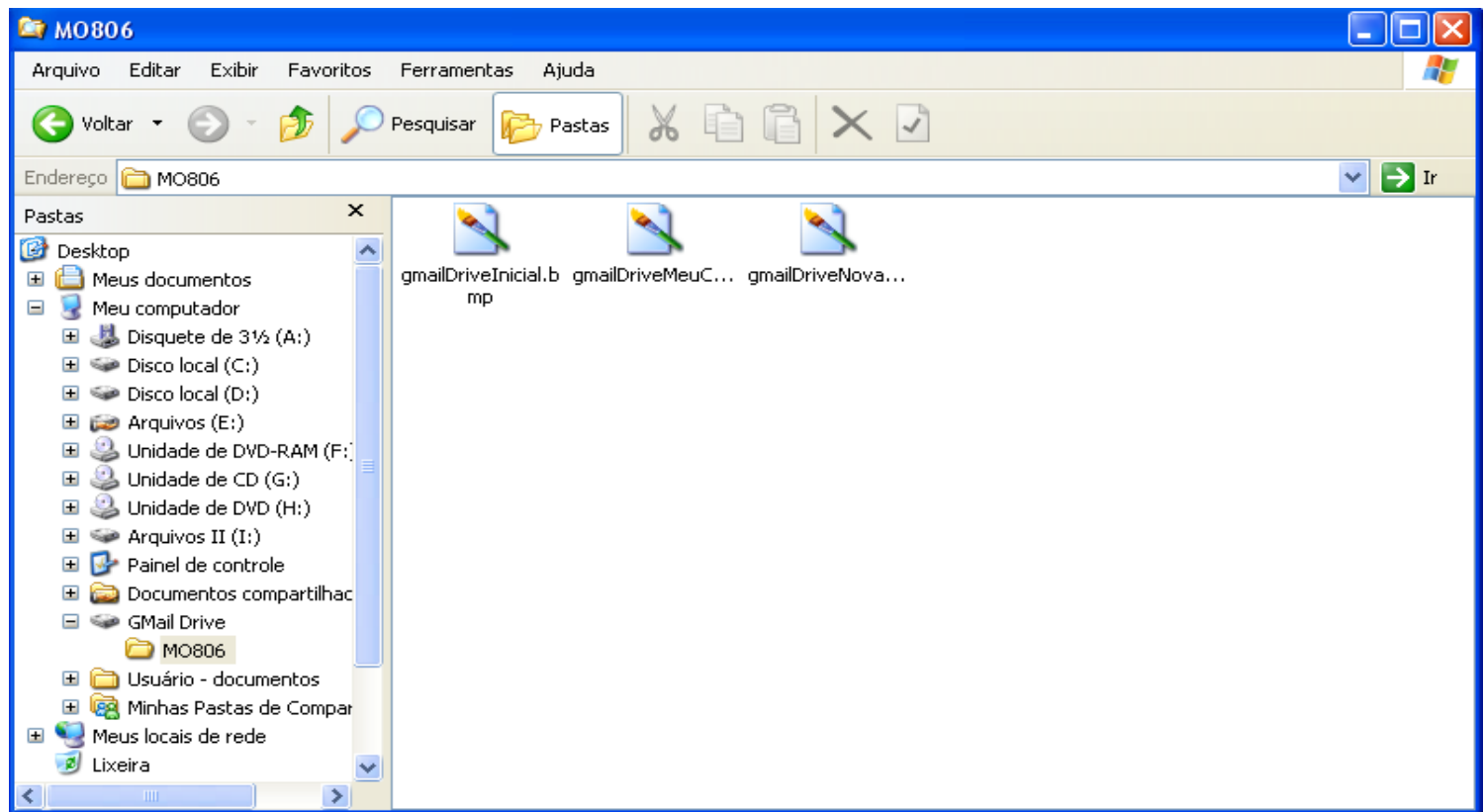
# Gmail File System - Windows

- Transferência:



# Gmail File System - Windows

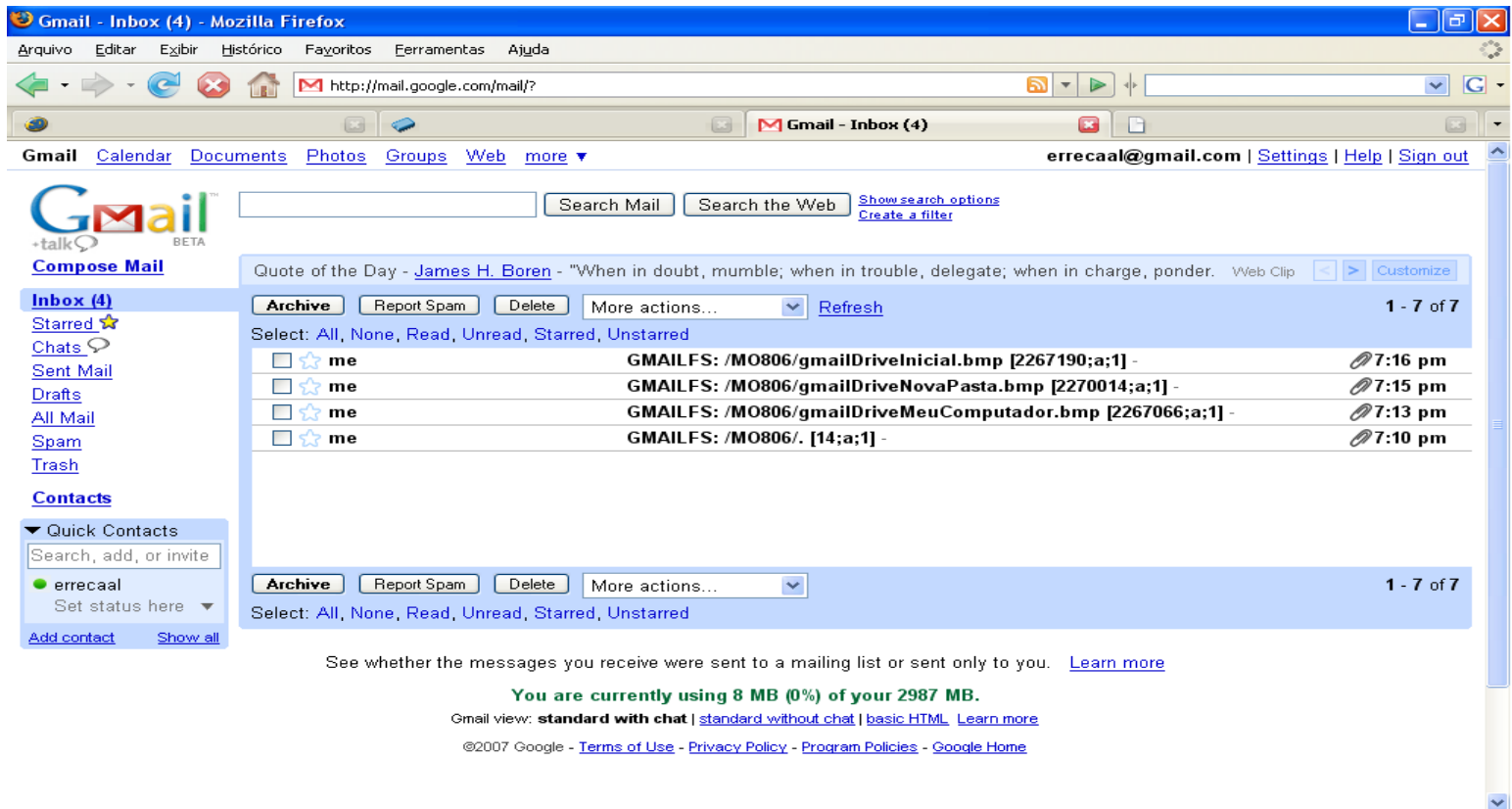
- Pasta modificada:





# Gmail File System - Windows

- E-mail:



The screenshot shows the Gmail web interface in a Mozilla Firefox browser window. The browser's address bar displays the URL <http://mail.google.com/mail/>. The Gmail interface includes a top navigation bar with links for [Gmail](#), [Calendar](#), [Documents](#), [Photos](#), [Groups](#), [Web](#), and [more](#). The user's email address, [errecaal@gmail.com](#), is visible in the top right corner, along with links for [Settings](#), [Help](#), and [Sign out](#).

The main content area shows the "Inbox (4)" with a search bar and buttons for "Search Mail" and "Search the Web". Below the search bar, there is a "Quote of the Day" by James H. Boren. The inbox list displays four email items, each with a star icon, a checkbox, and a preview of the subject line:

Star	Checkbox	Subject Line	Time
<input type="checkbox"/>	<input checked="" type="checkbox"/>	GMAILFS: /M0806/gmailDriveInicial.bmp [2267190;a;1] -	7:16 pm
<input type="checkbox"/>	<input checked="" type="checkbox"/>	GMAILFS: /M0806/gmailDriveNovaPasta.bmp [2270014;a;1] -	7:15 pm
<input type="checkbox"/>	<input checked="" type="checkbox"/>	GMAILFS: /M0806/gmailDriveMeuComputador.bmp [2267066;a;1] -	7:13 pm
<input type="checkbox"/>	<input checked="" type="checkbox"/>	GMAILFS: /M0806/. [14;a;1] -	7:10 pm

At the bottom of the interface, there is a status bar indicating the user is currently using 8 MB (0%) of their 2987 MB storage. Links for [standard with chat](#), [standard without chat](#), [basic HTML](#), and [Learn more](#) are provided. The footer includes copyright information for 2007 Google and links to [Terms of Use](#), [Privacy Policy](#), [Program Policies](#), and [Google Home](#).

# Tópicos abordados

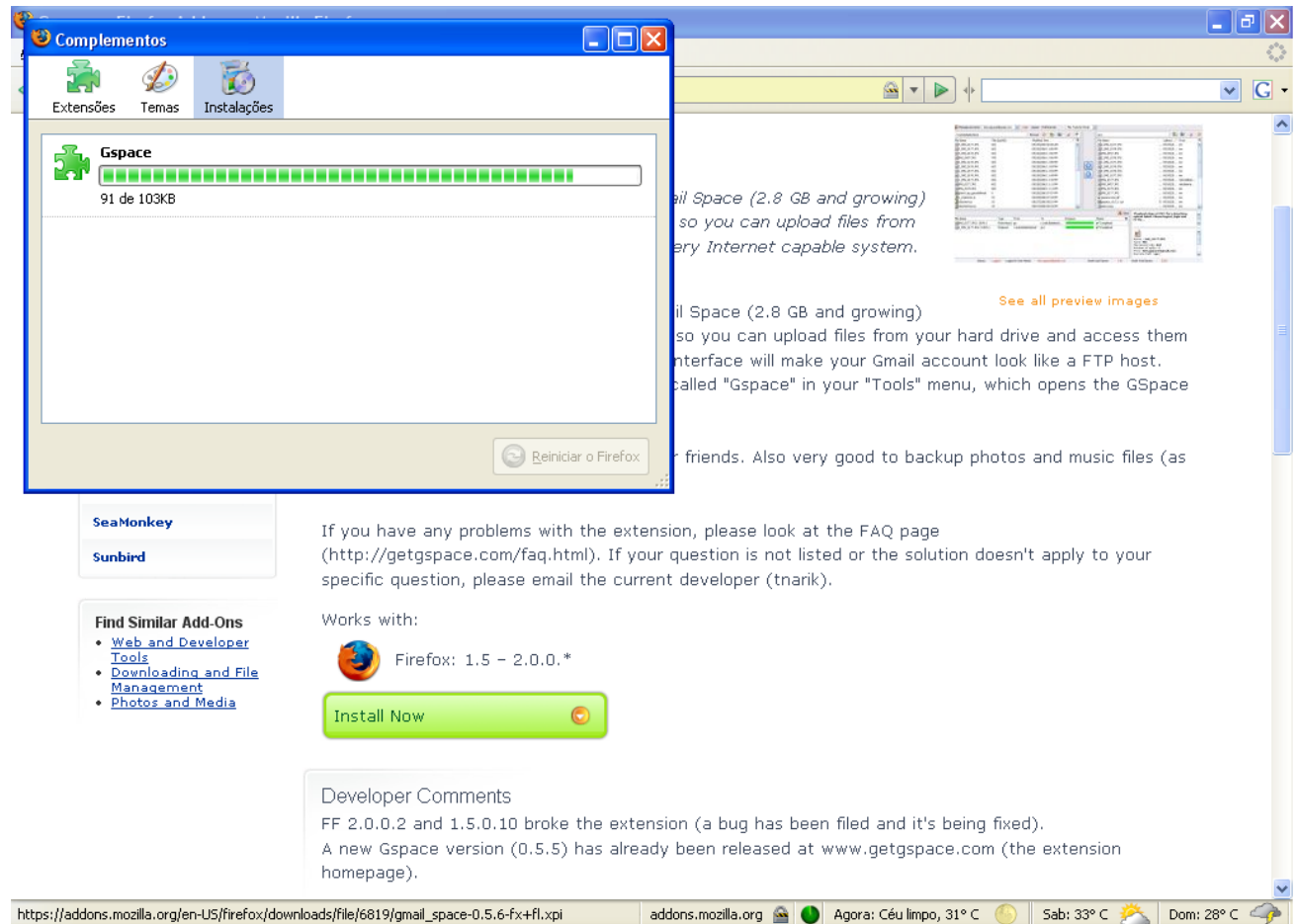
- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Gmail File System - Gspace

- É uma extensão do navegador Firefox que permite usar o Gmail Space para armazenamento de arquivo. Ele atua como um drive online, de tal maneira que é possível fazer upload de arquivos e os acessar de qualquer computador com Internet. A interface fará com que a conta do Gmail se pareça com um FTP host

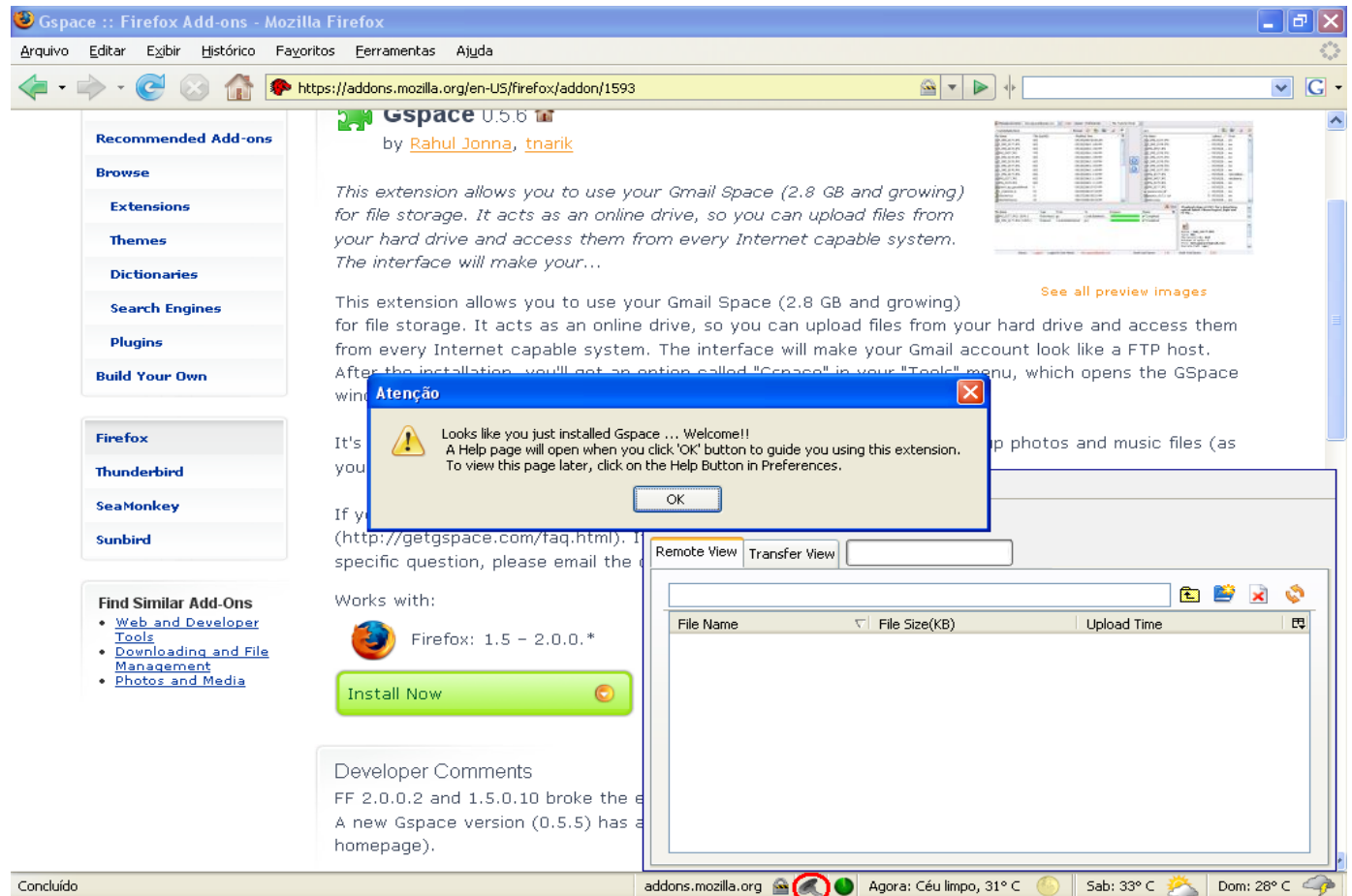
# Gmail File System - Gspace

- Instalação:



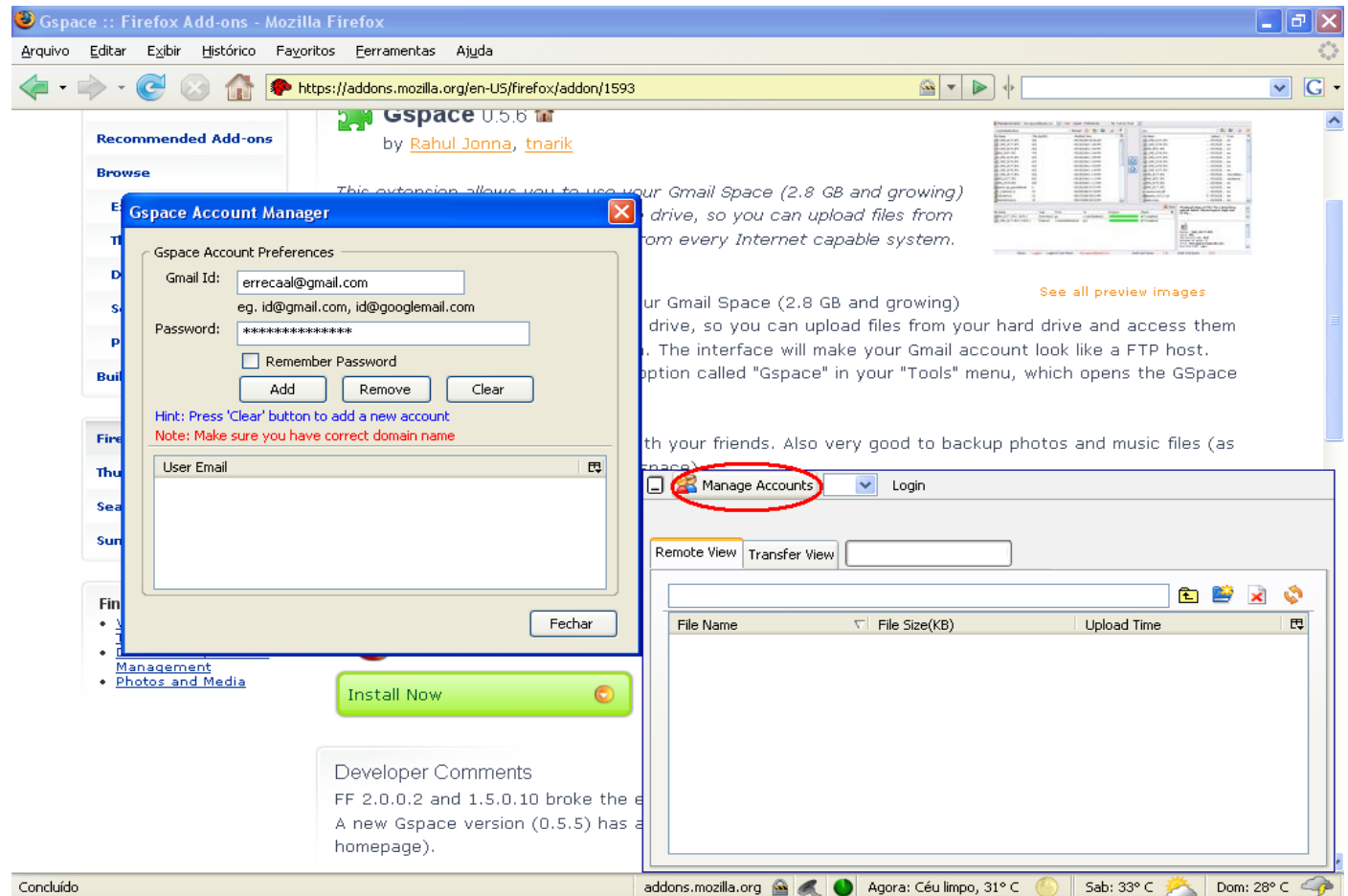
# Gmail File System - Gspace

- Início:



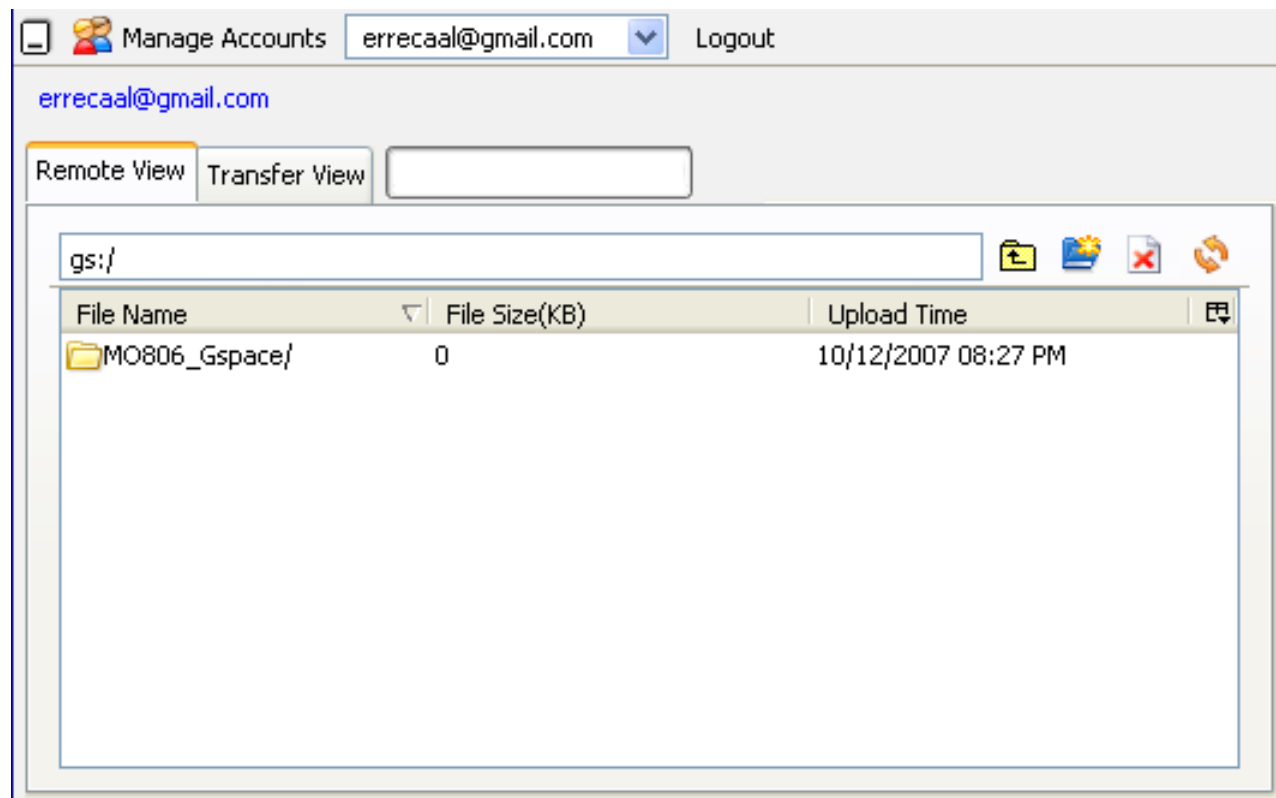
# Gmail File System - Gspace

- Login:



# Gmail File System - Gspace

- Nova pasta:



# Gmail File System - Gspace

- Transferência:

The screenshot displays the Gspace web interface within a browser window. The interface includes a sidebar with navigation links, a main content area with text and a file list, and a bottom section with developer comments. A red arrow points from the file list in the main content area to a detailed file transfer window on the right. This window shows the file 'gspaceInicio.bmp' with a progress bar at 11% and a status of 'Processing'.

Developer Comments

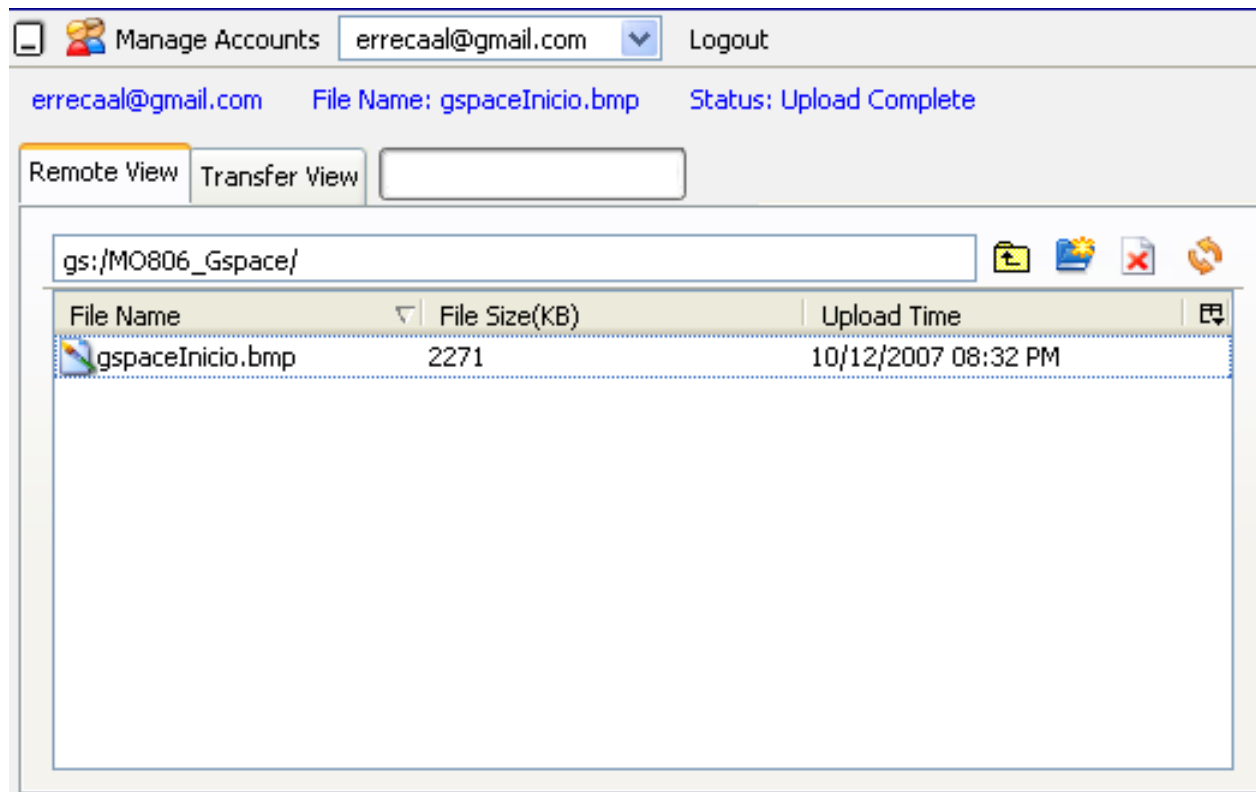
FF 2.0.0.2 and 1.5.0.10 broke the e  
A new Gspace version (0.5.5) has a  
homepage).

File Name	Progress	Status
gspaceInicio.bmp ( 11% )	<div></div>	Processing



# Gmail File System - Gspace

- Pasta modificada:



# Gmail File System - Gspace

- Inbox:

Gmail - Inbox (6) - Mozilla Firefox

Arquivo Editar Exibir Histórico Favoritos Ferramentas Ajuda

http://mail.google.com/mail/?

Gmail Calendar Documents Photos Groups Web more ▼ errecaal@gmail.com | Settings | Help | Sign out

**Gmail** **BETA**

[Compose Mail](#)

**Inbox (6)**

[Starred](#) [Chats](#) [Sent Mail](#) [Drafts](#) [All Mail](#) [Spam](#) [Trash](#) [Contacts](#)

Quick Contacts

Search, add, or invite

errecaal Loading...

Wired Top Stories - [Swiss Believe They've Found Foolproof Encryption for Voting Machines](#) - 11 hours ago Web Clip Customize

Archive Report Spam Delete More actions... Refresh 1 - 6 of 6

Select: All, None, Read, Unread, Starred, Unstarred

<input type="checkbox"/>	me	GSPACE gspacelnicio.bmp 1524 1 1 2271 gs:/MO806_Gspace/ d\$ - 2271 <metada 8:32 pm
<input type="checkbox"/>	me	GSPACE MO806_Gspace/ 4457 1 1 0 gs:/ d\$ - 0 <metadata><ST><![CDATA[10/12/ 8:27 pm
<input type="checkbox"/>	me	GMAILFS: /MO806/gmailDriveInicial.bmp [2267190;a;1] - 7:16 pm
<input type="checkbox"/>	me	GMAILFS: /MO806/gmailDriveNovaPasta.bmp [2270014;a;1] - 7:15 pm
<input type="checkbox"/>	me	GMAILFS: /MO806/gmailDriveMeuComputador.bmp [2267066;a;1] - 7:13 pm
<input type="checkbox"/>	me	GMAILFS: /MO806/. [14;a;1] - 7:10 pm

Archive Report Spam Delete More actions... 1 - 6 of 6

Select: All, None, Read, Unread, Starred, Unstarred

Automatically [forward](#) your Gmail messages to another email account. [Learn more](#)

**You are currently using 11 MB (0%) of your 2987 MB.**

Gmail view: [standard with chat](#) | [standard without chat](#) | [basic HTML](#) [Learn more](#)

©2007 Google - [Terms of Use](#) - [Privacy Policy](#) - [Program Policies](#) - [Google Home](#)

# Tópicos abordados

- Sistemas de arquivos
- Sistemas de arquivos distribuídos
- Google File System
- Gmail File System
  - Linux
  - Windows
  - Gspace
- Referências

# Referências

- Modern Operating Systems 2nd Edition – Andrew S. Tanenbaum
- Monografia Google File System - Gizelle Sandrini Lemos
- The Google File System – Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung

# Referências

- Gmail File System – Linux:
  - <http://richard.jones.name/google-hacks/gmail-filesyst>
  - <http://www.guiadohardware.net/artigos/gmail-filesyst>