

# 4 passos até à Gestão de Capacidade



[Pesquisa avançada](#)



## Pesquise e aceda a páginas do passado

Veja ou reveja páginas que já desapareceram.

São mais de 1000 milhões de conteúdos arquivados entre [1996](#) e [2011](#) que estão ao seu dispor.

[Conheça o projeto](#)

**FCCN**

Fundação para a Computação Científica Nacional  
Foundation for National Scientific Computing

# Estamos a perde a internet

- 50 dias - 50% dos documentos estão a ser alterados  
(Cho and Garcia-Molina. 2000)
- 1 ano - 80% dos documentos ficam inacessíveis  
(Ntoulas, Cho and Olson. 2004)
- 27 meses - 13% das referências na web desaparecem  
(<http://webcitation.org/>. 2007)

# Queremos evitar isto



## The page cannot be found

The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

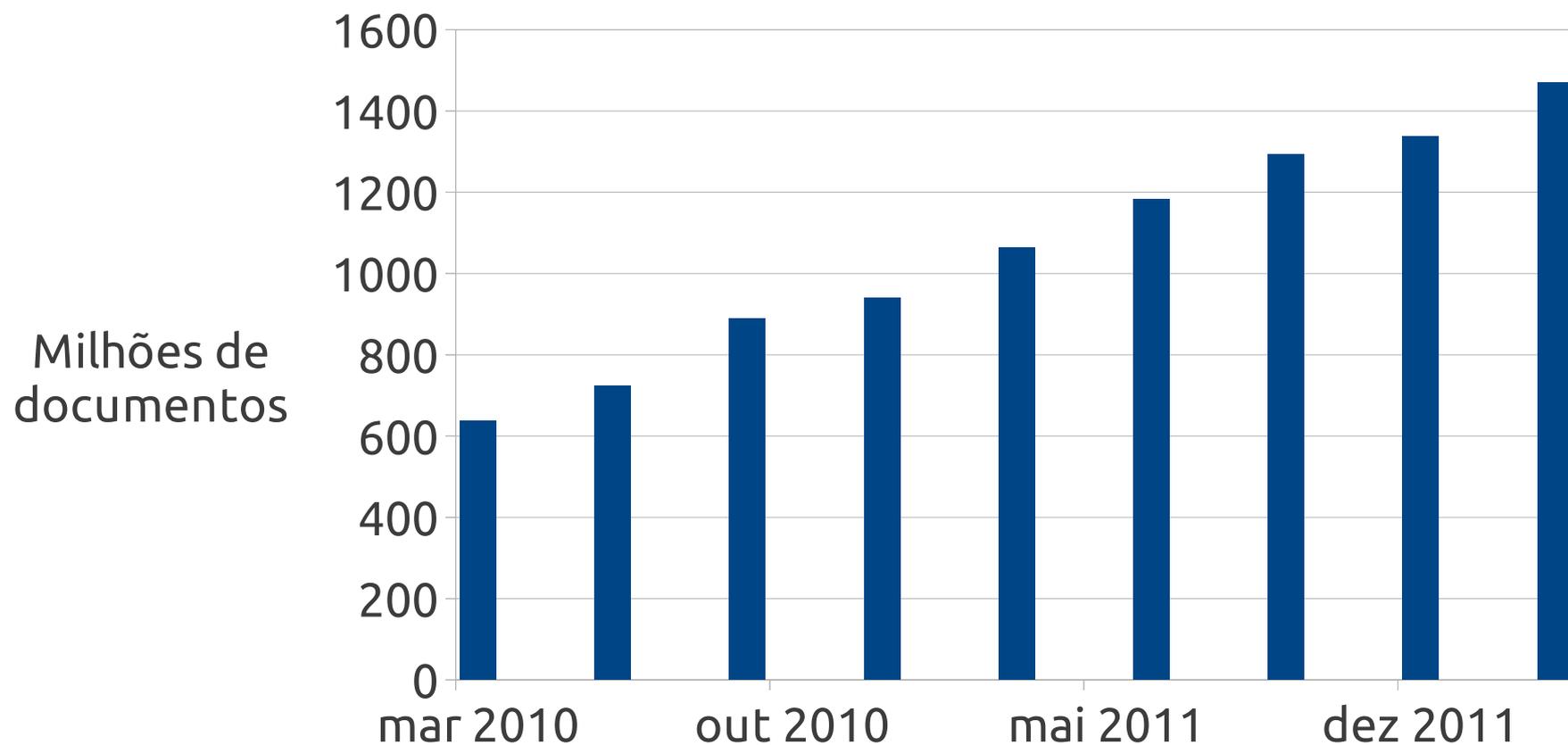
Please try the following:

- If you typed the page address in the Address bar, make sure that it is spelled correctly.
- Open the [httpd.apache.org](http://httpd.apache.org) home page, and then look for links to the information you want.
- Click the  [Back](#) button to try another link.
- Click  [Search](#) to look for information on the Internet.

HTTP 404 - File not found  
Internet Explorer

**404**  
**ERROR**

# 1500 Milhões de documentos (45 TB)



# Pesquisa por texto



sapo



Pesquisar no Arquivo

entre:

01/01/1996



e:

31/12/2011



[Pesquisa avançada](#)

Resultados 1 a 10 de 188.862.279

## [SAPO - Portugal Online!](#)

**18 Dezembro, 2009** - [outras datas](#)

SAPO - Portugal Online! Saltar para: Pesquisa [1], Lista de Serviços [2], Notícias [3] ou Destaques SAPO [4] SAPO.pt  
Pesquisa SAPO Web Imagens Notícias Blogs Produtos Directório PAi PBi Pesquisar ... SAPO Se está a ver este texto é porque o seu browser não está actualizado e não lhe permite ver este ...

<http://www.sapo.pt/>

## [SAPO - Servidor de Apontadores Portugueses](#)

**10 Dezembro, 1997** - [outras datas](#)

8a2 SAPO - Servidor de Apontadores Portugueses Ainda lhe restam dúvidas sobre o SAPO ? Esclareça-se! c4d  
Novidades Novos Links , Congressos , ... Ensino e Investigação Universidades , Institutos , Escolas , ...  
Comunicação Social Jornais , Rádios , Televisão , ... Entretenimento Desportos ...

<http://www.sapo.pt/>

## [Eu Não Desisto: abril 2004 Archives](#)

**17 Outubro, 2009** - [outras datas](#)

Jornal de Notícias, Minho, Braga, 17.12.2004, ou em [http://jn.sapo.pt/2004/12/17/minho ... .blogs.sapo.pt/arquivo/2004\\_04.html#128423](http://jn.sapo.pt/2004/12/17/minho...blogs.sapo.pt/arquivo/2004_04.html#128423) Posted by mauricio\_102 at 02:46 PM | Comentários: (20 ... Portugueses ... III". 30\_4\_04 - "LISTAGEM dos Artigos do Mês de Abril 2004". 28\_4\_04 - ""blogs.sapo" 25 ...

[http://eunaodesisto.blogs.sapo.pt/arquivo/2004\\_04.html](http://eunaodesisto.blogs.sapo.pt/arquivo/2004_04.html)

# Pesquisa por URL



ARQUIVO DA WEB PORTUGUESA

sapo.pt

entre: 01/01/1996 e: 31/12/2011

Pesquisar no Arquivo

[Pesquisa avançada](#)

Pretendia ver os resultados das páginas que contêm o texto: <http://sapo.pt?>

## Versões da página web guardadas no arquivo

Foram gravadas no arquivo 1.373 versões da página <http://sapo.pt> entre 1 Janeiro, 1996 e 27 Novembro, 2012.

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
	4	23	87	58	20	29	199	444	119	120	5	6	255	2
Jan	<a href="#">25 Jan</a>	<a href="#">29 Fev</a>	<a href="#">5 Jan</a>	<a href="#">24 Jan</a>	<a href="#">5 Fev</a>	<a href="#">16 Fev</a>	<a href="#">1 Jan</a>	<a href="#">1 Jan</a>	<a href="#">2 Jan</a>	<a href="#">1 Jan</a>	<a href="#">20 Mai</a>	<a href="#">26 Mar</a>	<a href="#">22 Jan</a>	
Jan	<a href="#">25 Jan</a>	<a href="#">29 Fev</a>	<a href="#">6 Jan</a>	<a href="#">6 Fev</a>	<a href="#">10 Fev</a>	<a href="#">19 Mar</a>	<a href="#">2 Jan</a>	<a href="#">1 Jan</a>	<a href="#">5 Jan</a>	<a href="#">14 Mar</a>	<a href="#">24 Jun</a>	<a href="#">1 Abr</a>	<a href="#">23 Jan</a>	
Fev	<a href="#">8 Fev</a>	<a href="#">29 Fev</a>	<a href="#">7 Jan</a>	<a href="#">30 Mar</a>	<a href="#">19 Fev</a>	<a href="#">5 Abr</a>	<a href="#">3 Jan</a>	<a href="#">2 Jan</a>	<a href="#">7 Jan</a>	<a href="#">14 Mar</a>	<a href="#">26 Set</a>	<a href="#">5 Abr</a>		
Fev	<a href="#">8 Fev</a>	<a href="#">29 Fev</a>	<a href="#">8 Jan</a>	<a href="#">1 Abr</a>	<a href="#">20 Fev</a>	<a href="#">20 Mai</a>	<a href="#">4 Jan</a>	<a href="#">2 Jan</a>	<a href="#">7 Jan</a>	<a href="#">22 Out</a>	<a href="#">26 Set</a>	<a href="#">8 Abr</a>		
	<a href="#">9 Fev</a>	<a href="#">1 Mar</a>	<a href="#">19 Jan</a>	<a href="#">29 Mai</a>	<a href="#">24 Mar</a>	<a href="#">3 Jun</a>	<a href="#">4 Jan</a>	<a href="#">5 Jan</a>	<a href="#">9 Jan</a>	<a href="#">22 Out</a>	<a href="#">18 Dez</a>	<a href="#">9 Abr</a>		
	<a href="#">20 Fev</a>	<a href="#">3 Mar</a>	<a href="#">24 Jan</a>	<a href="#">30 Mai</a>	<a href="#">12 Abr</a>	<a href="#">9 Jun</a>	<a href="#">5 Jan</a>	<a href="#">6 Jan</a>	<a href="#">11 Jan</a>		<a href="#">18 Dez</a>	<a href="#">12 Abr</a>		
	<a href="#">20 Fev</a>	<a href="#">3 Mar</a>	<a href="#">30 Jan</a>	<a href="#">4 Jun</a>	<a href="#">19 Abr</a>	<a href="#">9 Jun</a>	<a href="#">5 Jan</a>	<a href="#">10 Jan</a>	<a href="#">12 Jan</a>			<a href="#">13 Abr</a>		

# Em 2010...

- Chefe: Quanto tempo demora até ficarmos sem espaço em disco?
- Eu: ???

# Exemplo

- Páginas apresentadas ao usuário em menos 5 segundos
- Espaço em disco necessário para manter o serviço
- Número de consultas por segundo abaixo dos 500 ms

# Gestão de capacidade

- Saber que recursos devem ser adicionados, quando e porquê.
- **Definição:** Processo que define a quantidade de recursos necessários para garantir a continuidade do produto mediante alterações na procura.

# Não é velocidade



# Tens gasolina suficiente para chegar?



# Chegar a uma boa gestão de capacidade



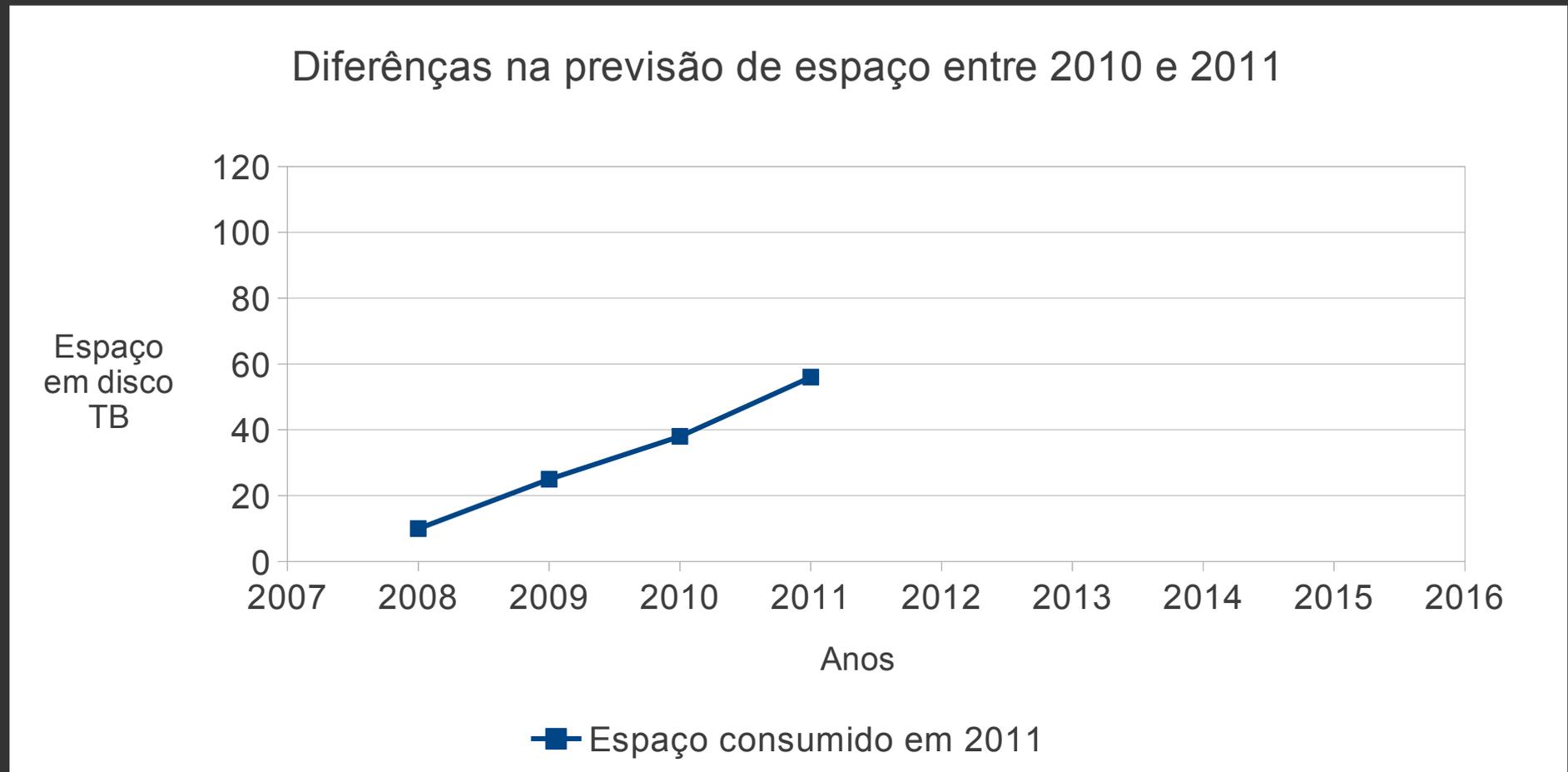
# 1. Medir a utilização dos recursos

- Precisamos de instrumentos de medição
- Qual o espaço em disco utilizado por ano?

13 TB por ano



# Consumo atual de espaço em disco por ano

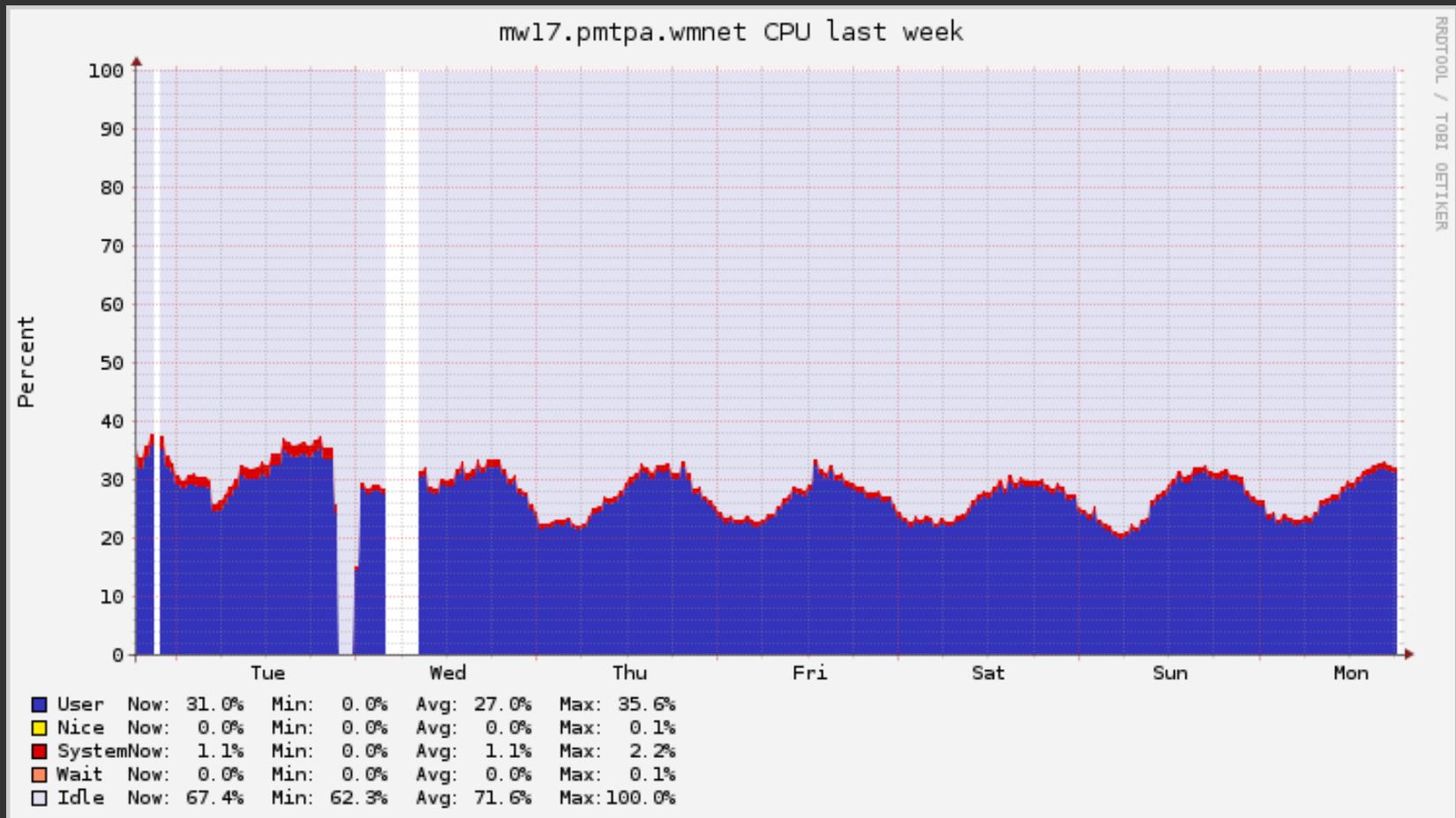


## 2. Descobrir os limites de capacidade

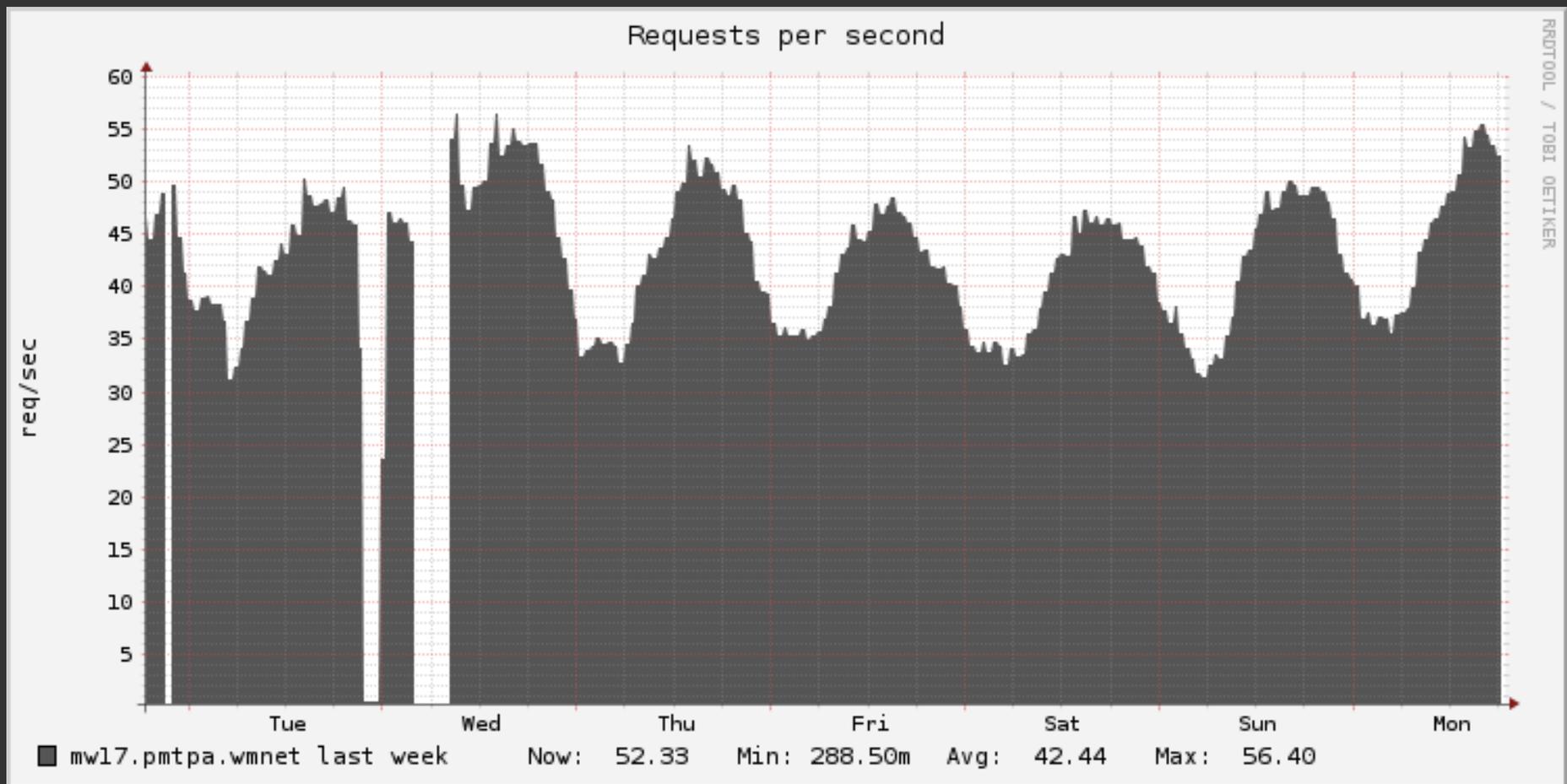
- Relacionar a utilização com os recursos disponíveis



# Recursos (CPU, Memória, Disk I/O, ...)



# Utilização

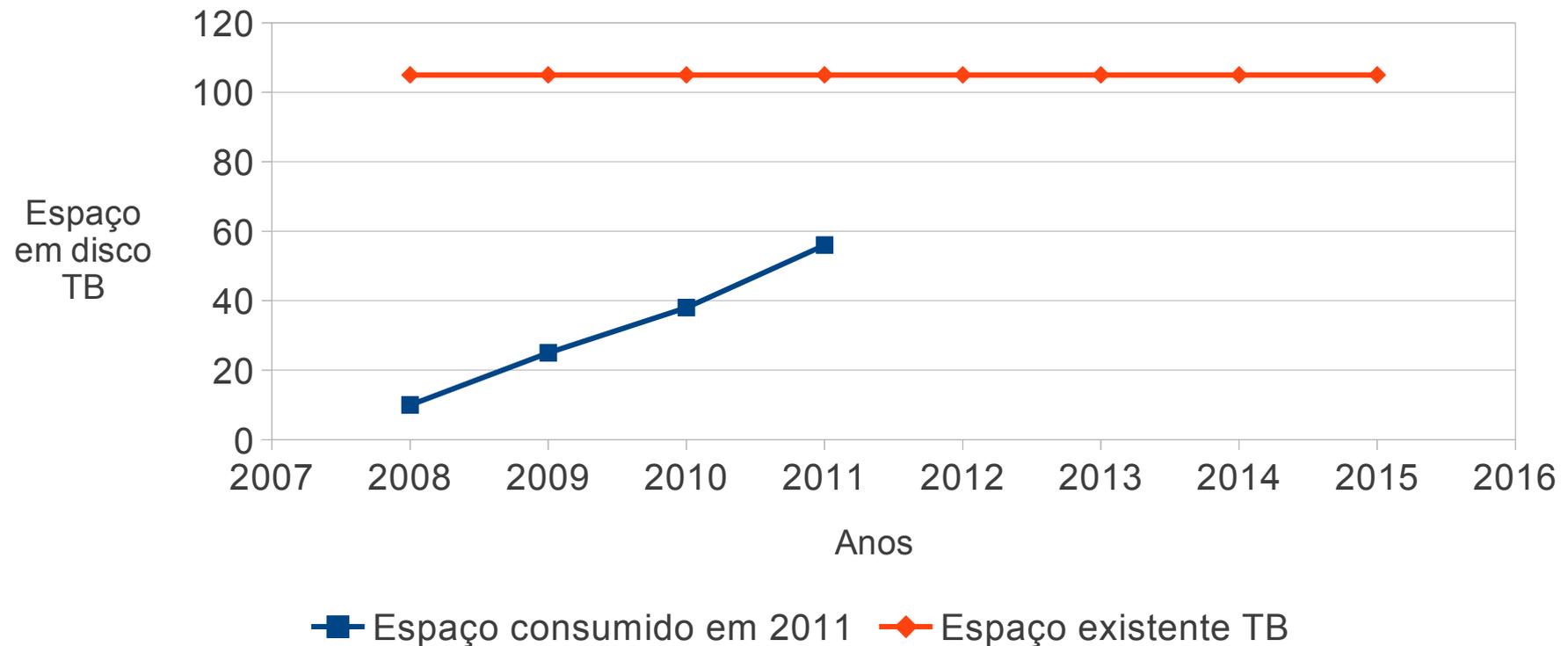


# Limites calculados em produção



# Tenho limite de utilização

Diferenças na previsão de espaço entre 2010 e 2011

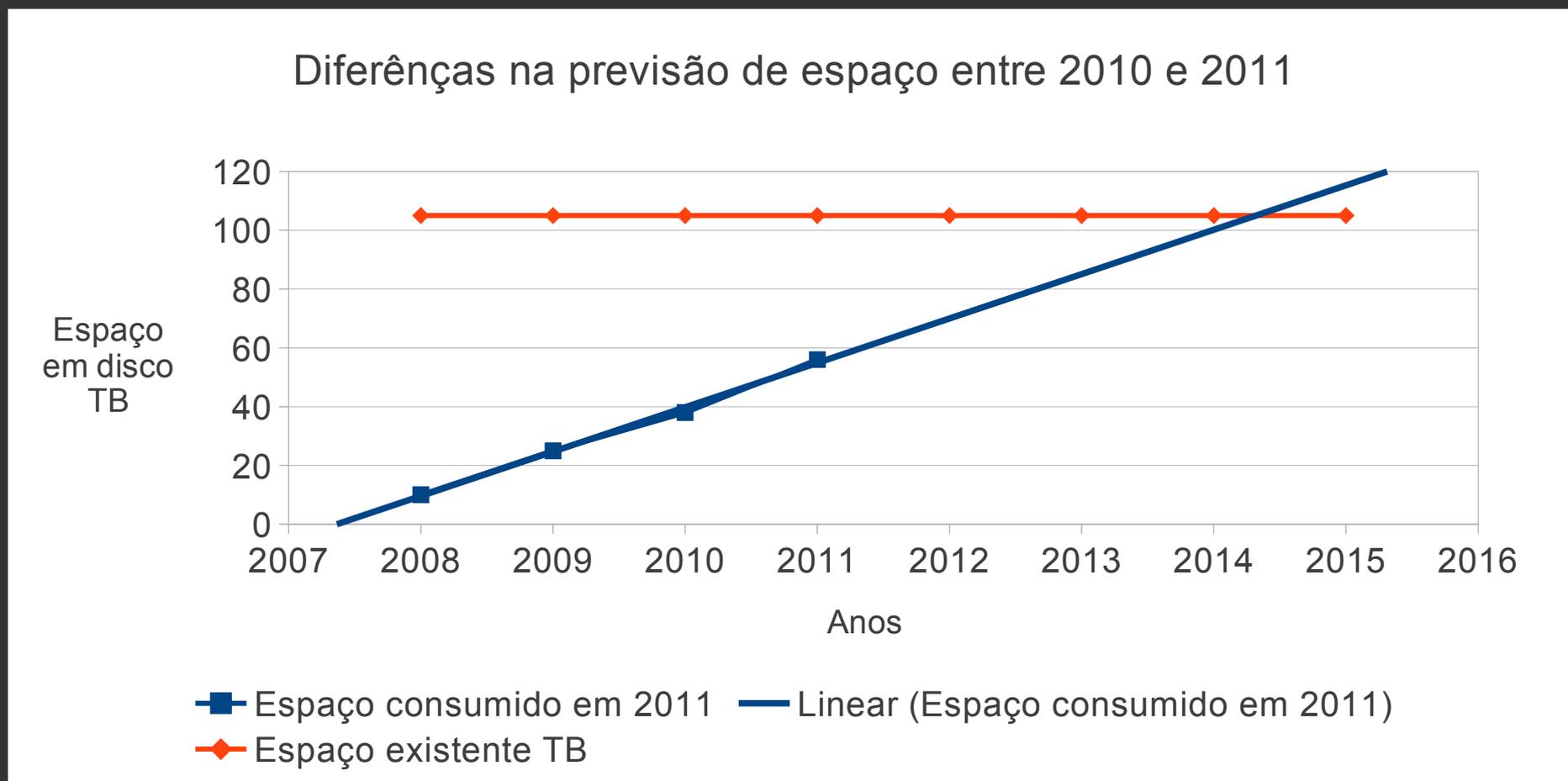


# Previsão

## previsão para os próximos dias

DOM	SEG	TER	QUA	QUI*
				
24° 15°	30° 15°	31° 17°	31° 18°	32° 17°

# Existirá espaço em disco até 2014



# 3. Adicionar recursos

#	Name	Q1, 2012	Q2, 2012		
		March	April	May	June
1	Descobrir necessidades	[Task Bar]			
2	Aprovar compras		[Task Bar]		
3	Consultar vendedores		[Task Bar]		
4	Entrega do equipamento			[Task Bar]	
5	Instalar equipamento				[Task Bar]
6					

# Passar para produção



ARQUIVO DA WEB PORTUGUESA

[Pesquisa avançada](#)

 **Pesquise e aceda a páginas do passado**  
Veja ou reveja páginas que já desapareceram.

São mais de 1000 milhões de conteúdos arquivados entre [1996](#) e [2011](#) que estão ao seu dispor.

[Conheça o projeto](#)

# Instalar manualmente

- Tempo de instalação do sistema operacional  
45 minutos
- Tempo de instalação de 10 sistemas operacionais:  
 $45 \text{ minutos} * 10 \text{ máquinas} = 7.5 \text{ horas}$
- Custo:  
 $30 \text{ reais / hora} * 7.5 \text{ horas} = 225 \text{ reais}$

# Temos de automatizar

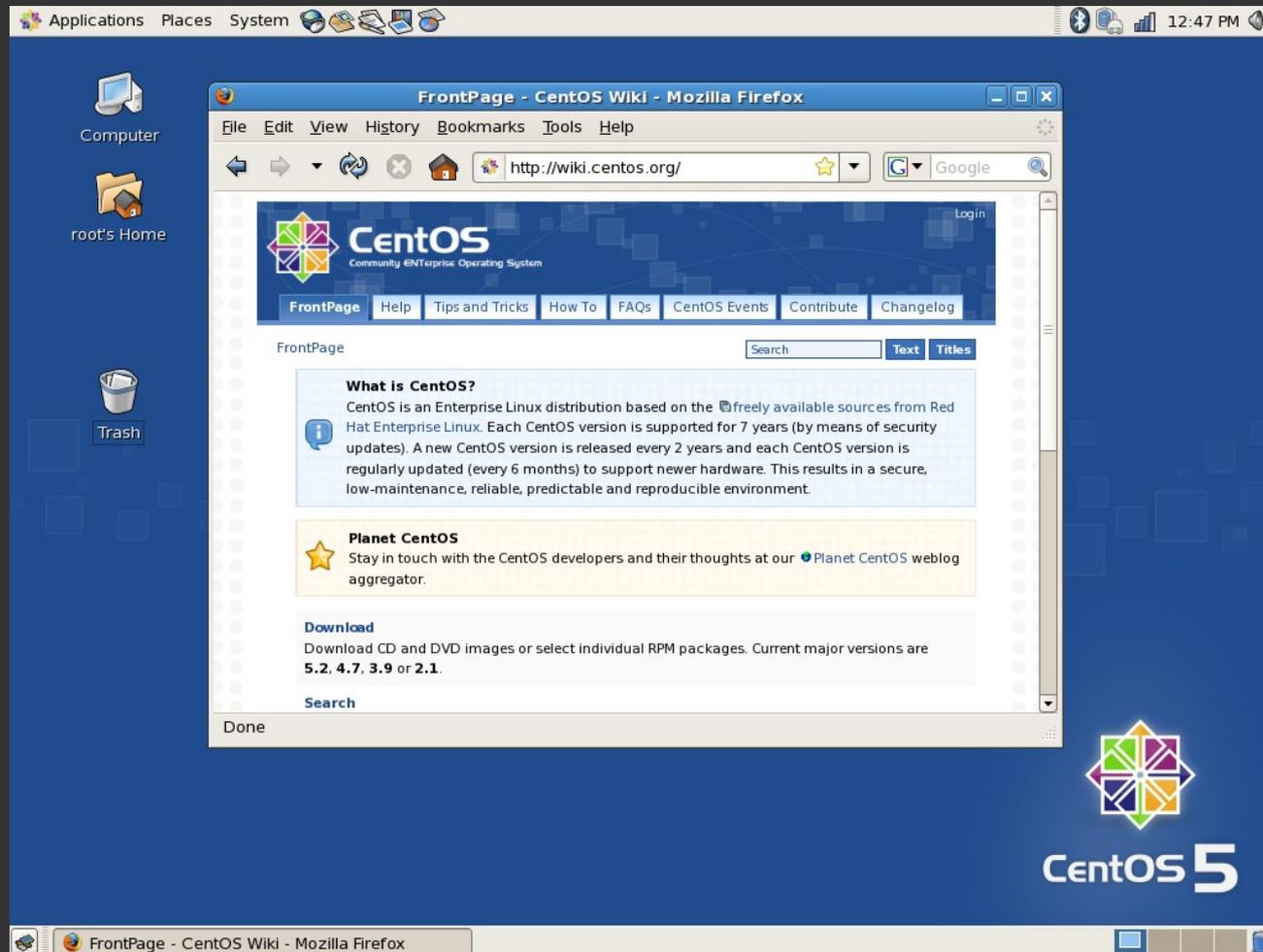
- Objectivo:

Reconstruir todo o serviço tendo apenas os computadores, o repositório de código e backups da aplicação.

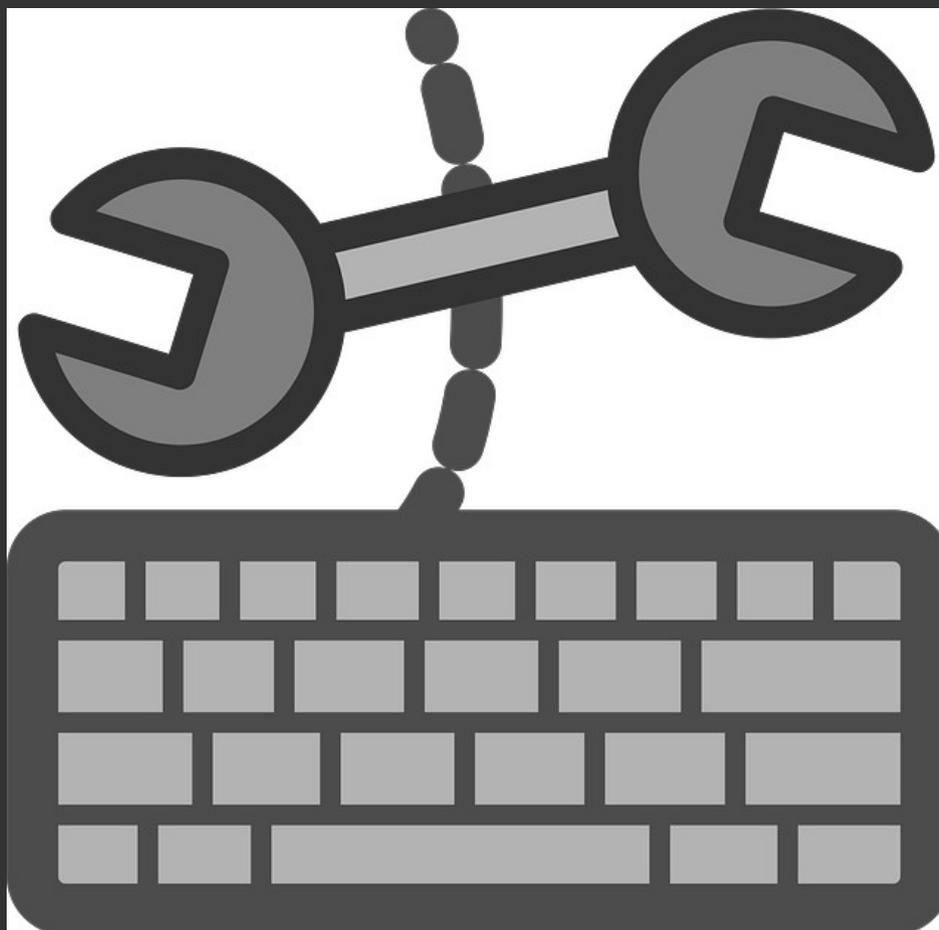
# Instalação automática de servidores

- Poupa dinheiro
- Minimizar o tempo até produção
- Mantém consistência no sistema

# Sistema operacional instalado



# Como configurar



# No começo

```
ssh root@web01.exemplo.com.br
```

```
# vim /etc/httpd/conf/httpd.conf
```

- Ligar cache
- Alterar email

# Passado 2 meses

```
ssh root@web02.exemplo.com.br
```

```
# vim /etc/httpd/conf/httpd.conf
```

```
– Alterar email
```

# Configuração Manual

- Prós
  - Funciona em 2/3 servidores
  - Fácil e rápido
- Contras
  - Configurações não replicáveis
  - Não é possível auditar
  - Não existe histórico das alterações
  - Não existe documentação
  - Não é possível reproduzir em 100 servidores rapidamente
  - Não podemos reconstruir servidores em caso de falha

# Criação de imagens

- Prós
  - Reduz tempo até produção de novos servidores
  - Rápido a criar grande número de servidores
- Contras
  - Sem validação do estado
  - Alterações de configurações leva à reconstrução de máquinas
  - Não é auditável
  - Sem informação de história
  - Sem documentação

# Gestão de configuração

- “Receitas” para configurar servidores

```
class ntp {  
    package { "ntp":  
        ensure => installed  
    }  
    service { "ntp":  
        ensure => running  
    }  
}
```

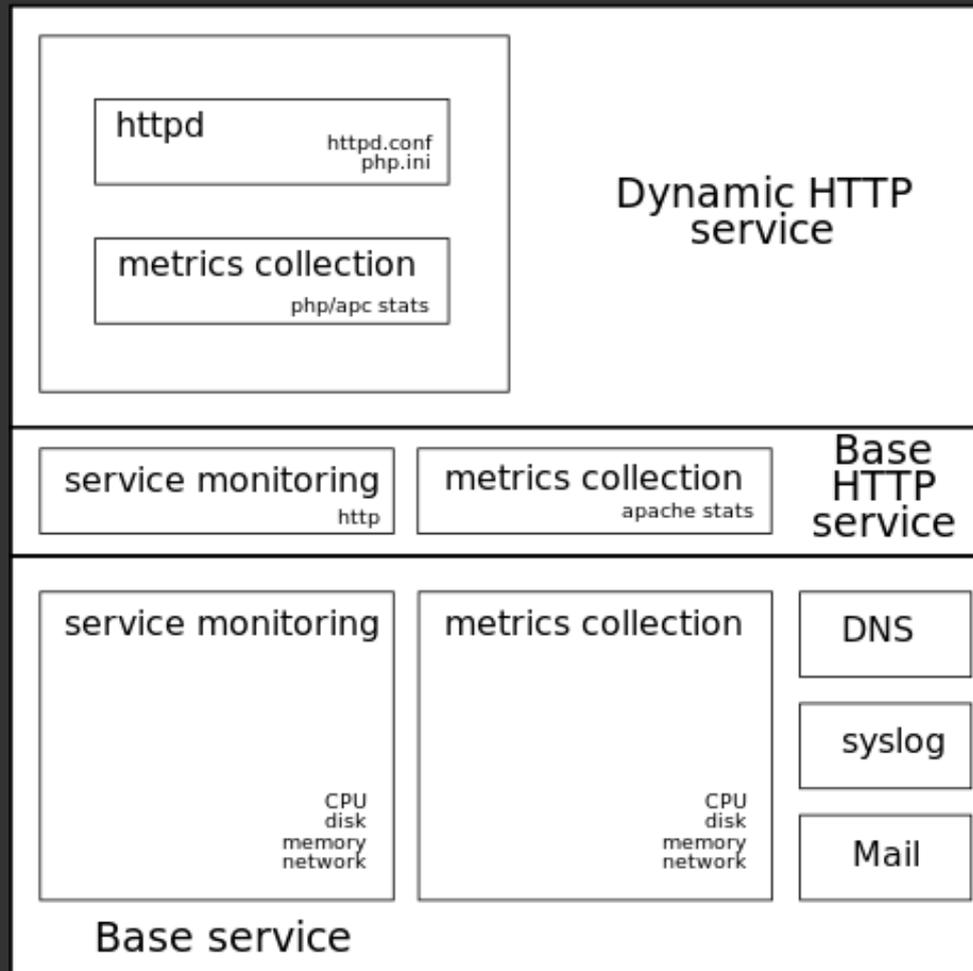
# Repositório de código e controlo de versões

- Tudo passa a ser código fonte
- Tudo será alterado em apenas um local
- Todos poderão ver qual é o código do sistema (configuração e código fonte do produto)

# Benefícios de gestão de configurações

- Reprodutibilidade
- Precisão
- Homogeneidade

# Configurar os tipos de máquinas



# Instalar o produto

experimental

Português English Ajuda

 ARQUIVO DA WEB  
PORTUGUESA

[Pesquisar no Arquivo](#)

[Pesquisa avançada](#)

 **Pesquise e aceda a páginas do passado**  
Veja ou reveja páginas que já desapareceram.

São mais de 1000 milhões de conteúdos arquivados entre [1996](#) e [2011](#) que estão ao seu dispor.

[Conheça o projeto](#)

Alguns momentos relevantes registados no arquivo  
Navegue por momentos que ficaram para a história.

 [José Saramago](#)

 [Expo '98](#)

# “1 clique” | Quem? Quando? O quê?

The screenshot shows the Jenkins web interface for a project named "03. Deploy to Arquivo.pt". The interface is organized into several sections:

- Navigation Menu (Left):** Includes links for "Back to Dashboard", "Status", "Changes", "Workspace", "Build Now", "Delete Project", "Configure", and "All Changes".
- Project Title:** "Project 03. Deploy to Arquivo.pt".
- Workspace:** A folder icon labeled "Workspace".
- Recent Changes:** A document icon labeled "Recent Changes".
- Build History (Left):** A table showing the last 8 builds with their status (green for success, red for failure) and timestamps.

Build #	Timestamp
#29	Jul 16, 2012 3:00:01 PM
#28	Mar 30, 2012 7:01:31 PM
#27	Mar 28, 2012 4:41:13 PM
#26	Mar 20, 2012 11:55:02 AM
#25	Mar 16, 2012 3:29:29 PM
#24	Jan 11, 2012 3:53:50 PM
#23	Jan 11, 2012 3:34:44 PM
#22	Jan 11, 2012 3:25:08 PM
- Downstream Projects:** A section titled "Downstream Projects" containing a link for "04. Test deploy in Production (arquivo.pt)".
- Permalinks:** A section titled "Permalinks" containing a list of links for specific build events:
  - Last build (#29), 4 mo 10 days ago
  - Last stable build (#29), 4 mo 10 days ago
  - Last successful build (#29), 4 mo 10 days ago
  - Last failed build (#22), 10 mo ago
  - Last unsuccessful build (#27), 8 mo 0 days ago

# Sistema instalado, recursos em utilização



The screenshot shows the homepage of the Portuguese Web Archive. In the top left corner, there is a blue diagonal banner with the word "experimental" in white. The top right corner features language options: "Português", "English", and "Ajuda". The main header consists of a logo with three stacked horizontal bars (blue, green, blue) and the text "ARQUIVO DE WEB PORTUGUESA". Below the logo is a search bar with a blue "Pesquisar no Arquivo" button and a "Pesquisa avançada" link. A large green checkmark is overlaid on the search bar area. Below the search bar, there is a section titled "Pesquise e descubra páginas do passado" with a sub-header "Veja ou reveja páginas que desapareceram." and a paragraph stating "São mais de 1000 milhões de conteúdos arquivados entre 1996 e 2011" with a link "Conheça o projeto". At the bottom, there is a section titled "Alguns momentos relevantes registados no arquivo" with the sub-header "Navegue por momentos que ficaram para a história." and two preview cards: one for "José Saramago" and another for "Expo '98".

# Processo demora 6 meses

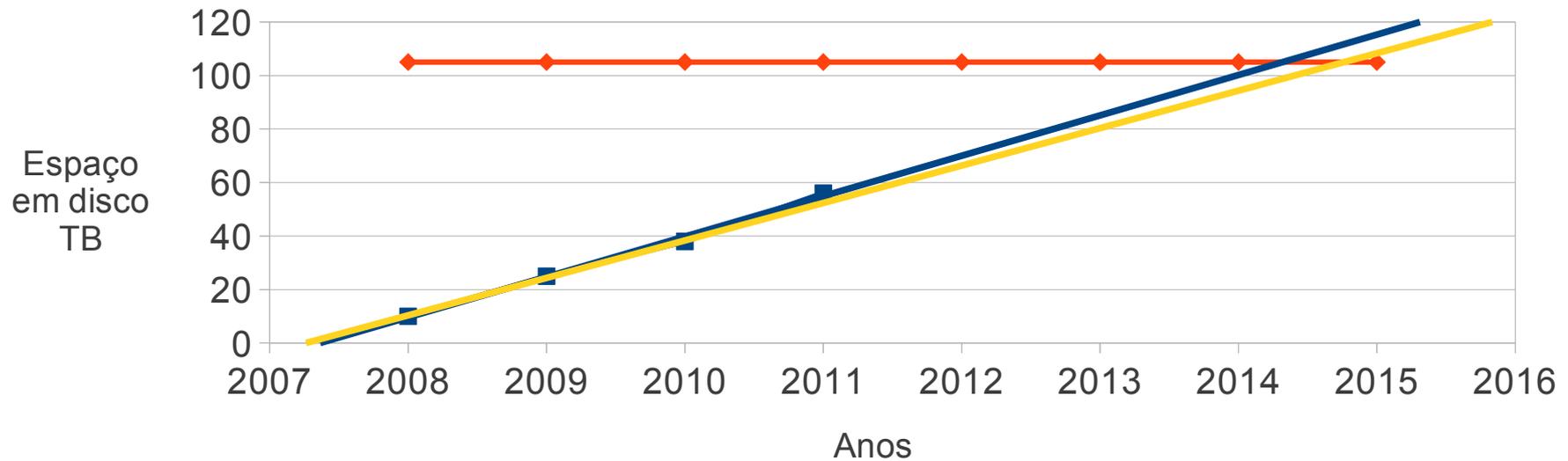


## 4. Repetir o processo

- Já adicionamos novos recursos.
- Repetir o processo, com estes novos recursos quando ficaremos novamente sem espaço em disco?

# 4. Repetir o processo

Diferenças na previsão de espaço entre 2010 e 2011

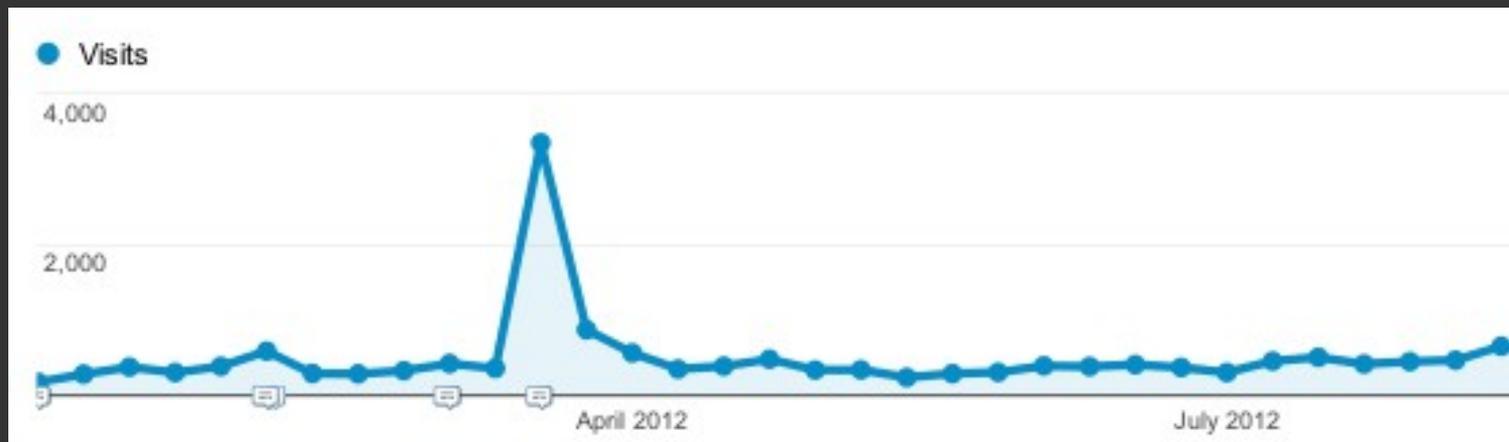


- Espaço consumido em 2010
- Linear (Espaço consumido em 2010)
- Espaço consumido em 2011
- Linear (Espaço consumido em 2011)
- Espaço existente TB

# Gestão de capacidade

- Medir utilização de recursos
  - Ter um sistema de medição dos recursos completo
- Descobrir limites
  - Ligar métricas com recursos
- Adicionar Recursos
  - Utilizar automatismos
- Repetir o processo
  - Recalcular para verificar os resultados

# Tráfego espontâneo



# Notícia num site importante



# Desabilitar funções pesadas

Q *encontre na globo.com*

☀️ 30° 16° São Paulo, SP ▾

[notícias ▾](#) | [esportes ▾](#) | [entretenimento ▾](#) | [tecnologia ▾](#) | [vídeos ▾](#) | [email ▾](#) | [assine já](#) | [central](#)

 **CASO ELIZA SAMUDIO** Fernanda começa a depor daqui a pouco [siga em tempo real >](#)

**Macarrão muda sua versão e surpreende ao incriminar Bruno**

- SIGA: Defesa de Bola diz que Macarrão o livrou

 **Bruno ainda não sabe de relato, diz advogado**

- Macarrão: 'Não sou gay, fui humilhado na prisão'
- Goleiro ganha lugar na calçada da fama do Fla

 **Beijo de viúva marca adeus a morto por copo**

 **Americana que foi agredida no Rio festeja alta**

**Confederações: acabam entradas**

**SUPERCLÁSSICO**

 **ASSISTA AGORA** 01:20

BI NA BOMBONERA  
**Fred aproveita chute cruzado e**

# Desabilitar funções pesadas

☀️ 30° 16° São Paulo, SP ▾

[notícias ▾](#) | [esportes ▾](#) | [entretenimento ▾](#) | [tecnologia ▾](#) | [vídeos ▾](#) | [email ▾](#) | [assine já](#) | [central](#)



**CASO ELIZA SAMUDIO** Fernanda começa a depor daqui a pouco [siga em tempo real >](#)

## Macarrão muda sua versão e surpreende ao incriminar Bruno

- SIGA: Defesa de Bola diz que Macarrão o livrou



**Bruno ainda não sabe de relato, diz advogado**

- Macarrão: 'Não sou gay, fui humilhado na prisão'
- Goleiro ganha lugar na calçada da fama do Fla



**Beijo de viúva marca adeus a morto por copo**



**Americana que foi agredida no Rio festeja alta**

### Confederações: acabam entradas

**SUPERCLÁSSICO**



**ASSISTA AGORA** 01:20

BI NA BOMBONERA  
**Fred aproveita chute cruzado e**

# Desabilitar funções pesadas

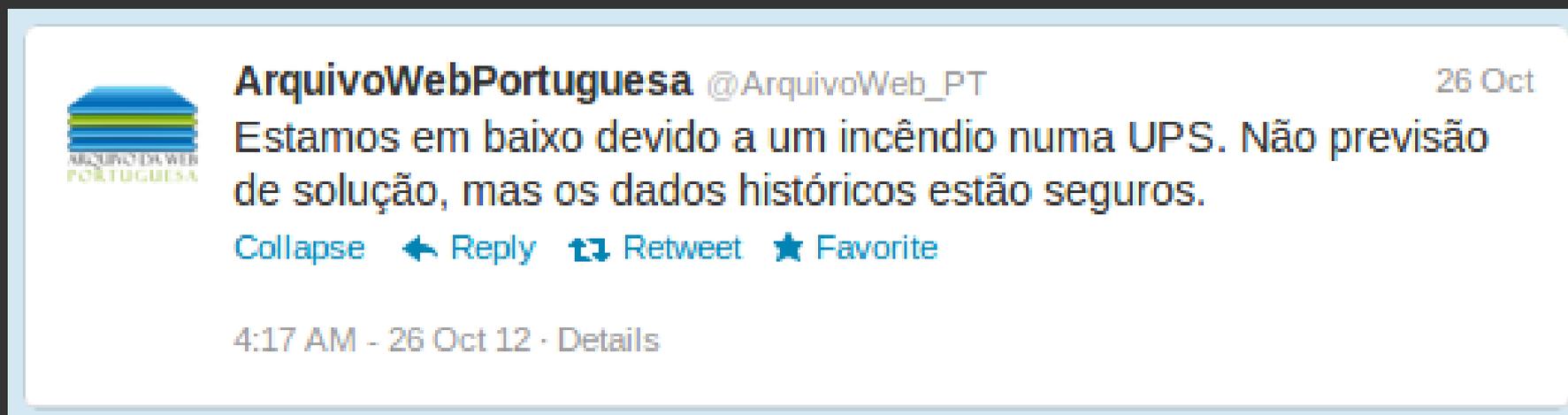
The screenshot shows the top section of the Globo.com website. At the top right, there is a search bar with the placeholder text "encontre na globo.com" and a "buscar" button. To the right of the search bar, the weather for São Paulo, SP is displayed as 30° 16° with a sun and cloud icon. Below the search bar, there is a horizontal navigation menu with links for "notícias", "esportes", "entretenimento", "tecnologia", and "vídeos", each with a dropdown arrow. To the right of these links are "email", "assine já", and "central" links. Below the navigation menu, there is a featured article banner for "CASO ELIZA SAMUDIO" with a sub-headline "Fernanda começa a depor daqui a pouco" and a "siga em tempo real" link. The main content area is divided into several sections. On the left, there is a large article titled "Macarrão muda sua versão e surpreende ao incriminar Bruno" with a sub-headline "SIGA: Defesa de Bola diz que Macarrão o livrou". Below this article is a smaller article titled "Bruno ainda não sabe de relato, diz advogado" with a sub-headline "Macarrão: 'Não sou gay, fui humilhado na prisão'" and "Goleiro ganha lugar na calçada da fama do Fla". In the center, there are two smaller articles: "Beijo de viúva marca adeus a morto por copo" and "Americana que foi agredida no Rio festeja alta". Below these is a section titled "Confederações: acabam entradas". On the right, there is a video player for a "SUPERCLÁSSICO" featuring Fred, with a "ASSISTA AGORA" button and a "01:20" duration indicator. Below the video player, there is a section titled "BI NA BOMBONERA" with a sub-headline "Fred aproveita chute cruzado e".

# Lidar com tráfego espontâneo

- Servir conteúdo estático
  - Processar site para transformar em estático
- Servir conteúdo de caches
  - Utilizar caches de páginas completas, exemplo: reverse-proxy
  - Utilizar *content delivery networks* CDN

# Manter canais de comunicação

- Ter canais de comunicação alojados fora da empresa
- Avisar utilizadores de quaisquer problemas, explicar o que se está a passar



# Conclusão

- Já lidamos com o crescimento tradicional do nosso site – mediante gestão de capacidade.
- Já conseguimos minimizar um crescimento espontâneo – notícia em site muito visitado.

Obrigado pela atenção

Tirem as vossas dúvidas.

Simão Fontes

[simaofontes@gmail.com](mailto:simaofontes@gmail.com)

<https://sites.google.com/site/simaofontes/>

# Fotos

- [flickr.com/photos/alohaowl/5145037507/](https://www.flickr.com/photos/alohaowl/5145037507/)
- [flickr.com/photos/peteredin/3174493833/](https://www.flickr.com/photos/peteredin/3174493833/)
- [flickr.com/photos/bambooly/298632541](https://www.flickr.com/photos/bambooly/298632541)
- [flickr.com/photos/ooohoooh/276100445/](https://www.flickr.com/photos/ooohoooh/276100445/)
- [flickr.com/photos/hpatton/6238331202/](https://www.flickr.com/photos/hpatton/6238331202/)
- [flickr.com/photos/ac4lt/4927460547/](https://www.flickr.com/photos/ac4lt/4927460547/)
- [flickr.com/photos/redjar/360113468/](https://www.flickr.com/photos/redjar/360113468/)
- [flickr.com/photos/jamescridland/613445810/](https://www.flickr.com/photos/jamescridland/613445810/)