

Análise Comparativa entre o BRF e Métodos Tradicionais para Gerenciamento de *Backups*

Tássia Camões Araújo

Projeto Final de Graduação 2006.2
Bacharelado em Ciência da Computação
IM-UFBA

13 de dezembro de 2006

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Sumário

Introdução

Conceitos Importantes

BRF - Backup, Remember, and Forget

Metodologia dos Experimentos

Evolução do BRF

Análise Comparativa

Conclusão

Links

Projeto Final
2006.2

Tássia Camões
Araújo

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Backup de dados:

cópia de segurança de um conjunto de dados que pode ser usada em caso de perda ou corrupção dos dados originais

- ▶ Política de *backups* de uma organização
periodicidade, mídias, automatização, etc
- ▶ Gerenciamento do espaço
- ▶ Ferramentas para gerência de *Backups*
Amanda, Bacula, BRF, etc

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Objetivos do Trabalho

Projeto Final
2006.2

Tássia Camões
Araújo

- ▶ Apresentar o BRF e sua proposta para gerenciamento de espaço
- ▶ Realizar análise comparativa entre políticas adotadas com BRF e Bacula

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Conceitos Importantes

Projeto Final
2006.2

Tássia Camões
Araújo

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

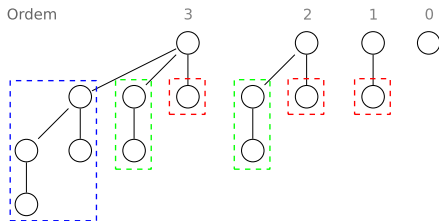
Conclusão

Links

- ▶ Tipos de *Backup*:
 - ▶ Completo
 - ▶ Diferencial
 - ▶ Incremental
- ▶ Necessidade de Rotação (reuso) de dispositivos
- ▶ Rotação "Avô, pai, filho"
- ▶ Persistência de dados

Árvores Binomiais:

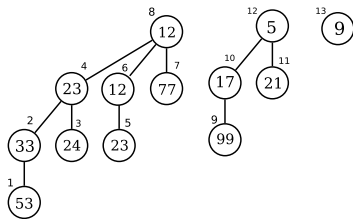
- ▶ Árvores ordenadas com filhos em ordem decrescente de grau
- ▶ Definição recursiva:
 1. Ordem 0: tem somente 1 nó
 2. Ordem k : uma árvore binomial de ordem $k - 1$ anexada à raiz de outra árvore binomial de ordem $k - 1$



Estruturas de Dados (2)

Heap Binomial:

- ▶ Conjunto de árvores binomiais (uma floresta)
- ▶ Todas as árvores binomiais da heap são heaps de mínimo ou todas são heaps de máximo
- ▶ Contém no máximo uma árvore binomial para cada ordem



53	33	24	23	23	12	77	12	99	17	21	5	9
1	2	3	4	5	6	7	8	9	10	11	12	13

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

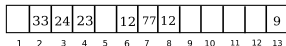
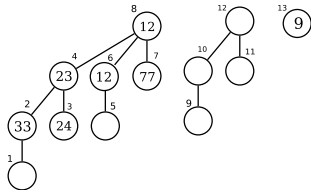
Conclusão

Links

Estruturas de Dados (3)

Heap Binomial com Buracos:

- ▶ Existência de vértices sem chave
- ▶ Todo nó das árvores binomiais é menor ou igual a seus descendentes



BRF - Backup, Remember, and Forget

Projeto Final
2006.2

Tássia Camões
Araújo

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

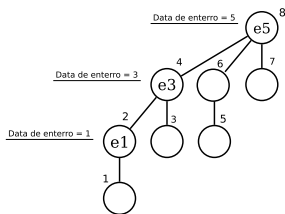
Conclusão

Links

- ▶ *Software Livre* escrito por Imre Simon em 2002
- ▶ Características-chave:
 - ▶ Otimiza espaço através de mecanismo de esquecimento de dados
 - ▶ Garante que a versão mais recente de todo arquivo descartado ou substituído está sempre disponível para recuperação
 - ▶ A implementação atual permite a reconstrução total de $2 + \log_2(n)$ estados anteriores, após n etapas.

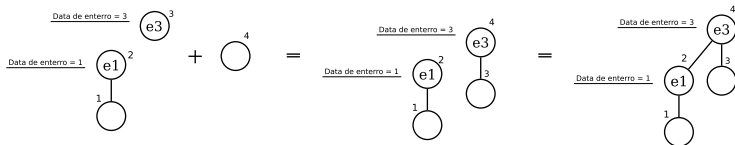
Backup, Remember, and Forget (1)

- ▶ Modelo semelhante ao *espelho + incrementais reversos*
- ▶ Datas de nascimento e de enterro dos eventos
- ▶ Eventos enterrados são guardados numa heap binomial de máximo com buracos
 - ▶ Rótulos dos vértices: etapa de criação dos vértices
 - ▶ Conteúdo dos vértices: eventos enterrados
 - ▶ Chave para comparação: etapa de enterro do eventos
 - ▶ Existe uma heap para cada conjunto de eventos de mesmo nome



Backup, Remember, and Forget (2)

- ▶ A cada etapa um novo vértice contendo a versão que acabou de sair do *Backup* é unido à heap da etapa anterior
- ▶ BRF garante que a versão mais recente de qualquer arquivo da heap estará sempre em R
- ▶ Promoção de eventos
- ▶ Buracos na heap entre eventos que competem



Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

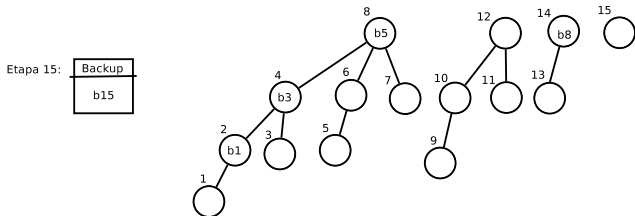
Conclusão

Links

Backup, Remember, and Forget (3)

Eventos da heap podem estar em:

- ▶ *Backup*
- ▶ *Remember*
- ▶ *Forget*



Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

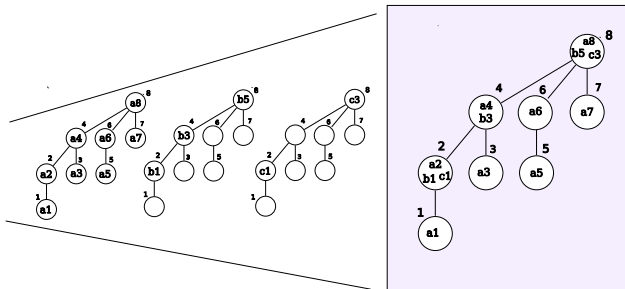
Análise
Comparativa

Conclusão

Links

Backup, Remember, and Forget (4)

Projeção das heaps do BRF:



Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Reconstrução da fotografia n :

- ▶ *Backup* atual é restaurado
- ▶ Volumes incrementais entre a raiz e n são aplicados em ordem decrescente
- ▶ Arquivos excedentes são apagados

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Implementação

Projeto Final
2006.2

Tássia Camões
Araújo

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Este complexo esquema tem implementação eficiente escrita em *Perl* e baseada no *rsync*

- ▶ Conceito de política de backups restringido a:
 1. Sistema utilizado
 2. Método de rotação entre dispositivos
 3. Persistência dos dados
- ▶ Dados provenientes de um sistema de grande porte monitorado durante 618 dias
- ▶ Desenvolvimento de simuladores em *Python*
- ▶ Realização de experimentos em duas fases

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Comportamento do BRF (1)

Evolução de espaço ocupado e quantidade de arquivos

Projeto Final
2006.2

Tássia Camões
Araújo

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

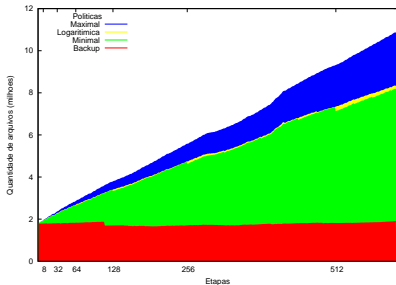
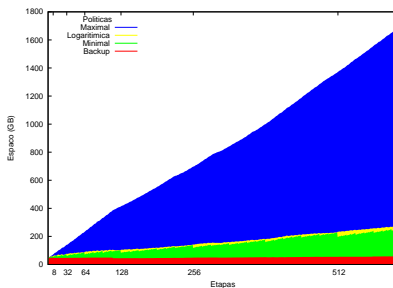
Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links



- ▶ *E-mail e cache*
 - ▶ *E-mail e cache com maildir*
 - ▶ *E-mail e cache com mbox*
- ▶ Arquivos de uso geral no sistema (que não são *e-mail* nem *cache*)
 - ▶ Dados de três usuários muito caros em termos de *Remember*
 - ▶ Arquivos de uso geral e de usuários com comportamento usual

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Parâmetro de comparação: demanda de espaço

- ▶ Resultado geral (58.06GB):
 - ▶ Minimal: 4.91 / Logaritmica: 5.34 / Maximal: 33.34
- ▶ Melhor resultado para arquivos de uso geral e de usuários com comportamento usual (43.50GB):
 - ▶ Minimal: 2.73 / Logaritmica: 2.85 / Maximal: 5.15
- ▶ Demanda de espaço na política logarítmica é perfeitamente aceitável, considerando a quantidade de reconstruções e as disponibilidade *eterna* das ultimas versões

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Parâmetros para a Comparação

Projeto Final
2006.2

Tássia Camões
Araújo

- ▶ Dispersão média dos estados reconstituíveis (em amplitude)
- ▶ Reconstrução relativa média
- ▶ Demanda de espaço relativa ao tamanho do backup

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

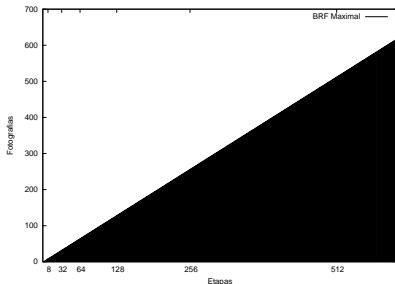
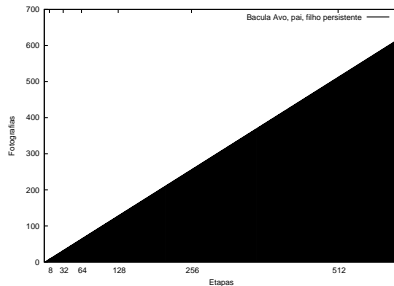
Links

Estados Reconstituíveis (1)

Políticas Maximais ou Persistentes

Projeto Final
2006.2

Tássia Camões
Araújo



Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

Resultados Obtidos

Projeto Final
2006.2

Tássia Camões
Araújo

Política	Precisão	Amplitude Média de Dispersão	Reconstrução Relativa Média	Qtd. Rec. Final	Demanda	
					Dados Com- pletos	Dados Seleci- onados
BRF Maximal	1 dia	309	100%	618	33.34	5.15
Bacula Avo, pai, filho		309	100%	618	55.65	26.58
BRF Lo- garítmico	$2^{(k+1)} - 2^k$, para estados em $[2^k, 2^{(k+1)}]$	149	6.3%	11	5.34	2.85
BRF Lo- garítmico persistente mensal	$2^{(k+1)} - 2^k$, para estados em $[2^k, 2^{(k+1)}]$ e no máximo de 1 mês	279	7.8%	25	13.03	5.98
BRF Avô, pai, filho persistente mensal	1 dia na última se- mana, 1 semana no último mês e 1 mês nos meses anterio- res	279	9.8%	33	13.15	5.98
Bacula Avô, pai, filho per- sistente mensal		308	9.5%	24	20.40	20.14
BRF Avô, pai, filho não persis- tente	1 dia na última se- mana, 1 semana no último mês e 1 mês no penúltimo mês	87	8.6%	12	5.42	2.82
Bacula Avô, pai, filho não persistente		45	7.6%	6	2.35	2.12
BRF Minimal	$\lceil n - (\lfloor \log_2(n) \rfloor + 1), n \rceil$	80	4%	6	4.91	2.73

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

- ▶ Em geral o BRF consegue uma demanda de espaço inferior à do Bacula (ausência de redundância)
- ▶ Disponibilizar versões mais recentes dos eventos enterrados não se mostrou muito custoso
- ▶ Contribuição:
 - ▶ Produção do melhor documento disponível para entendimento do BRF até o momento
- ▶ Trabalhos futuros:
 - ▶ Investir na remodelagem e implementação do BRF
 - ▶ Investigar sobre aspectos teóricos

Introdução

Conceitos
Importantes

BRF - Backup,
Remember, and
Forget

Metodologia dos
Experimentos

Evolução do BRF

Análise
Comparativa

Conclusão

Links

- ▶ Wiki do trabalho:

<http://conhecimento.incubadora.fapesp.br/portal/trabalhos/tassia/wiki/>

- ▶ Material adicional:

<http://marco.ime.usp.br/~is/brf-sim/>