RAMCloud Project

Bruno Padilha

19 de Junho de 2015

O que é o projeto RAMCloud ?

Sistema de armazenamento de dados de propósito geral que mantém os dados em memória RAM (DRAM) o tempo todo.

Objetivos: escalabilidade e baixa latência.

Motivação

➔ Os discos magnéticos foram (e ainda são) o principal modo de armazenamento de dados de sistemas on-line nos últimos 40 anos;

Motivação

- ➔ Os discos magnéticos foram (e ainda são) o principal modo de armazenamento de dados de sistemas on-line nos ultimos 40 anos;

- ➔ Sistemas de arquivos e SGBDs evoluiram para funcionar com HDs;

Motivação

- ➔ Os discos magnéticos foram (e ainda são) o principal modo de armazenamento de dados de sistemas on-line nos ultimos 40 anos;
- ➔ Sistemas de arquivos e SGBDs evoluiram para funcionar com HDs;
- ➔ A evolução do desempenho dos HDs não acompanhou a sua evolução de capacidade de armazenamento;

Motivação

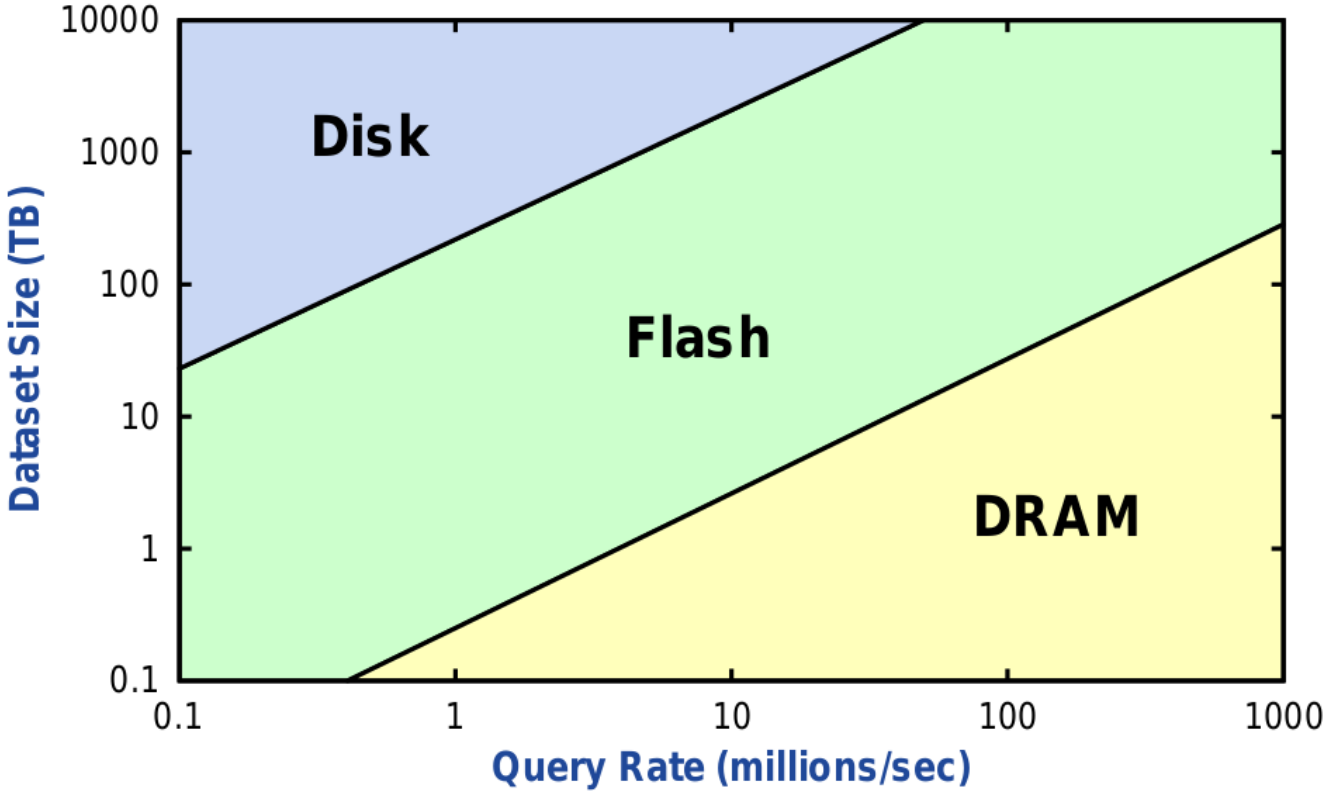
- ➔ Os discos magnéticos foram (e ainda são) o principal modo de armazenamento de dados de sistemas on-line nos ultimos 40 anos;
- ➔ Sistemas de arquivos e SGBDs evoluiram para funcionar com HDs;
- ➔ A evolução do desempenho dos HDs não acompanhou a sua evolução de capacidade de armazenamento;
- ➔ A dificuldade de escalabilidade da taxa de acesso aleatório aos dados acaba limitando a vazão de dados de um banco de dados relacional;

Evolução do disco rígido (HD)

Comparação entre os discos rígidos comuns disponíveis em 2009 e em 1985

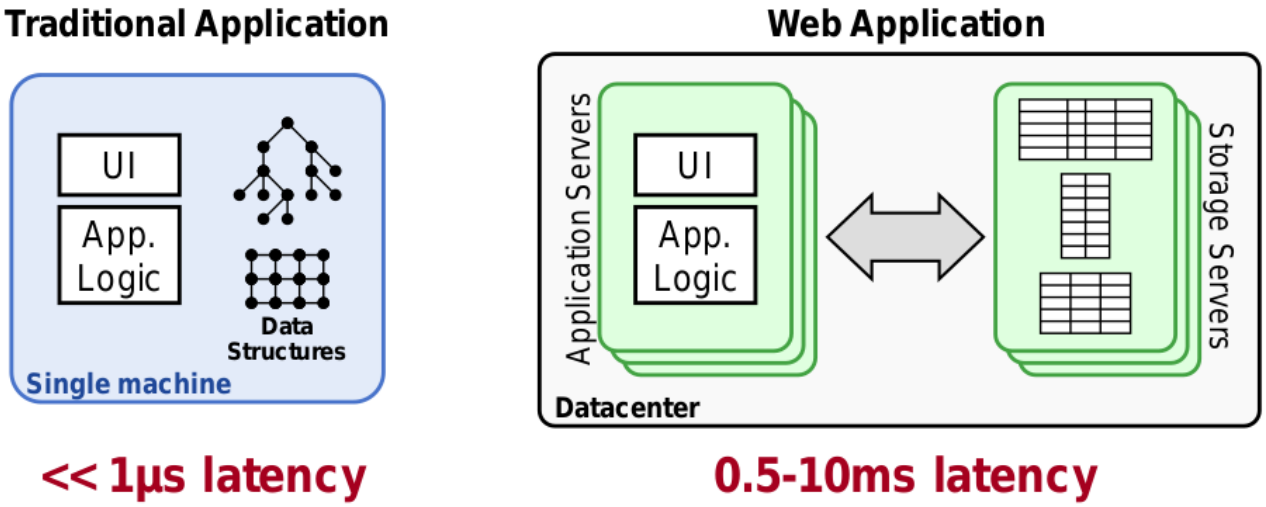
	Mid-1980's	2009	Change
Disk capacity	30 MB	500 GB	16667x
Max. transfer rate	2 MB/s	100 MB/s	50x
Latency (seek & rotate)	20 ms	10 ms	2x
Capacity/bandwidth (large blocks)	15 s	5000 s	333x
Capacity/bandwidth (1KB blocks)	600 s	58 days	8333x
Jim Gray's rule	5 min	30 hours	360x

Regra de Jim Gray



Por que a latência importa ?

Uma aplicação Web com dezenas de centenas de milhares de usuários não é capaz de ser executada em uma única máquina.



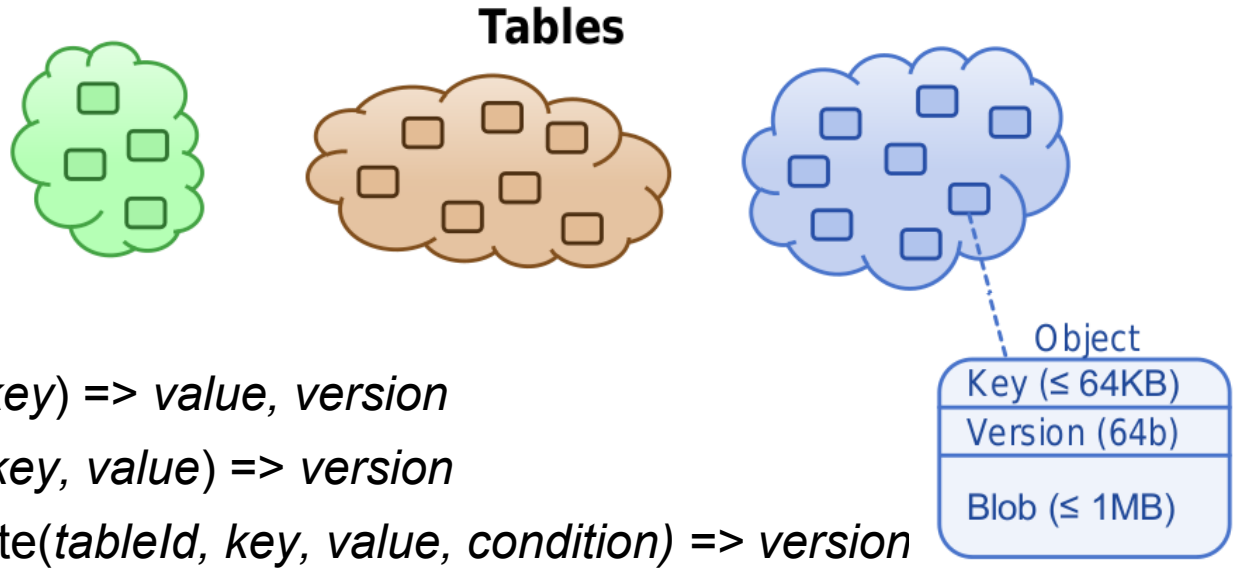
Possíveis soluções

- ➔ Novos sistemas de banco de dados (NoSQL);
- ➔ Caching;
- ➔ RAMCloud.

RAMCloud

- ➔ Armazenamento em DRAM para DataCenters (chave-valor);
- ➔ 1000 a 10000 (possivelmente 100000) servidores commodity;
- ➔ 64 a 256 GB de DRAM por servidor;
- ➔ Durabilidade e disponibilidade;
- ➔ Alta vazão (1M ops/sec/server);
- ➔ Baixa latência (5us pra leitura e 15us para escrita);
- ➔ Rápida recuperação de falhas (1-2s);
- ➔ Custo de cache volátil.

RAMCloud: Modelo de dados



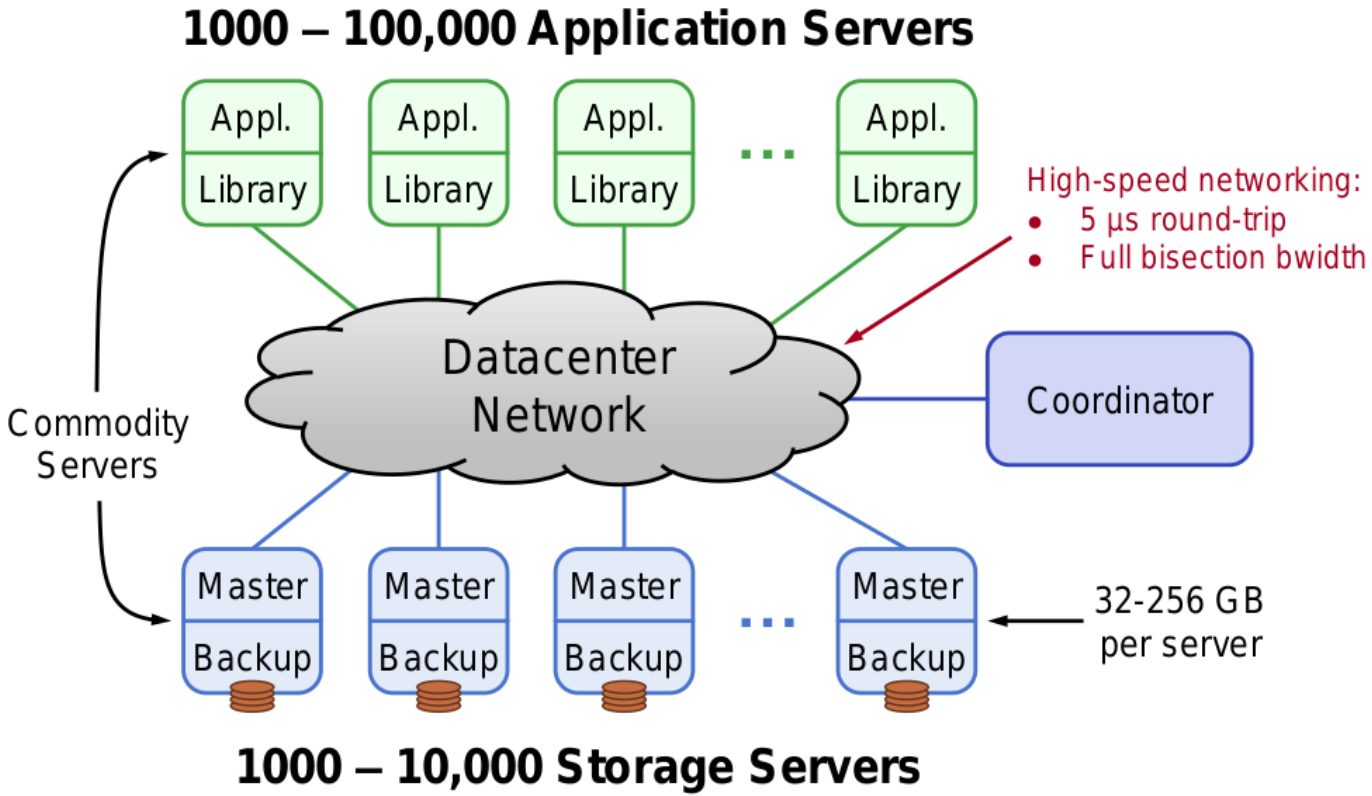
read(*tableId, key*) => *value, version*

write(*tableId, key, value*) => *version*

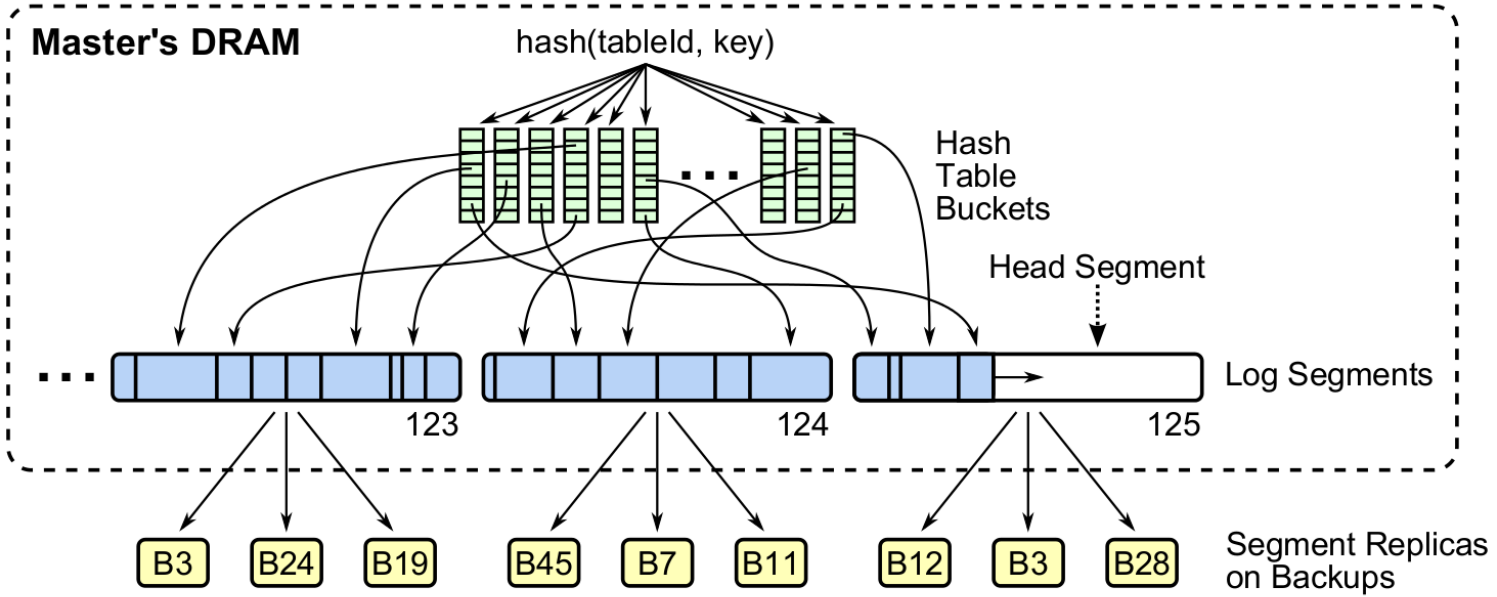
conditionalWrite(*tableId, key, value, condition)* => *version*

delete(*tableId,key*)

RAMCloud: Arquitetura

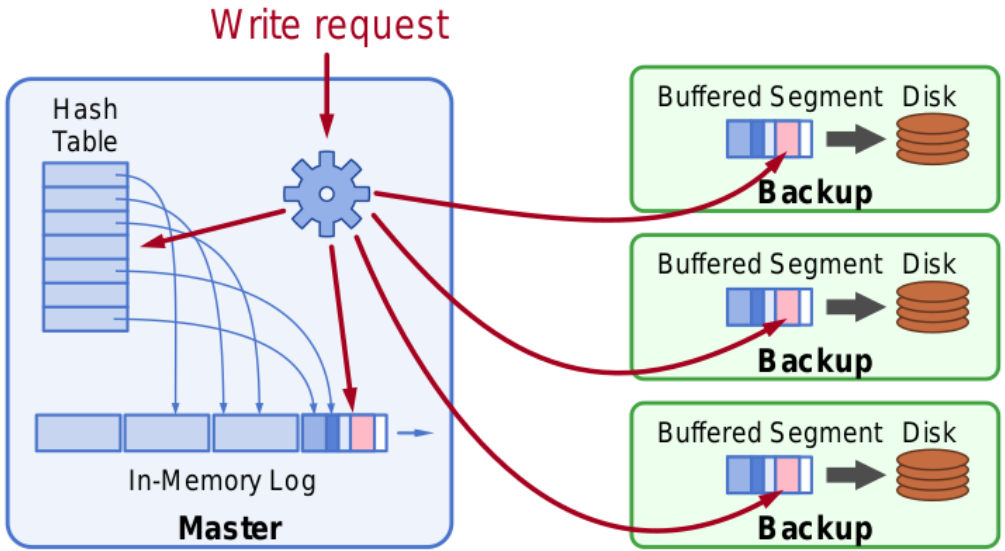


RAMCloud: Organização

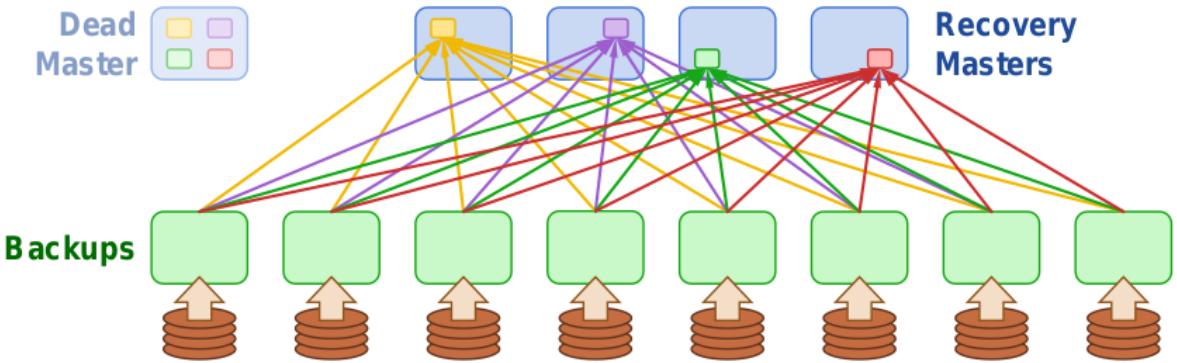


RAMCloud: Buffered logging

Backup escrito de modo assíncrono (sem espera por I/O).



RAMCloud: Crash recovery

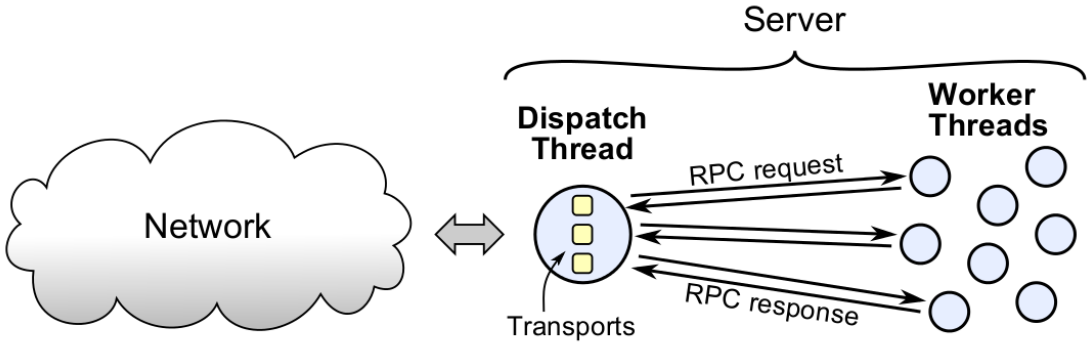


Os dados de um servidor mestre estão divididos em partições.

Em caso de falha, cada partição perdida é recuperada em um servidor distinto

RAMCloud: Latência

Component	Traversals	2009	Possible 2014	Limit
Network switches	10	100-300 µs	3-5 µs	0.2 µs
Operating system	4	40-60 µs	0 µs	0 µs
Network interface controller (NIC)	4	8-120 µs	2-4 µs	0.2 µs
Application/server software	3	1-2 µs	1-2 µs	1 µs
Propagation delay	2	1 µs	1 µs	1 µs
Total round-trip latency		**150-400µs**	**7-12 µs**	**2.4 µs**



RAMCloud: Utilizar

- ➔ Aplicações que utilizam muitos pedaços pequenos de dados independentes de uma dada requisição e precisam responder em tempo real;
- ➔ Aplicações que fazem muitas requisições dependentes que não podem ser feitas em paralelo (algoritmos em grafos). Exemplo: Facebook, Amazon, etc ...

RAMCloud: Não utilizar

- ➔ Aplicações que não fazem uso intensivo de dados;
- ➔ Aplicações que processam dados em lotes;
- ➔ Aplicações cujos dados dependem de propriedades de bancos de dados relacionais .

Referências

As seguintes referências encontram-se disponíveis na página do projeto RAMCloud Project:

https://ramcloud.atlassian.net/wiki/display/RAM/RAMCloud

- ➔ The RAMCloud Storage System;
- ➔ The Case for RAMCLoud;
- ➔ RAMCloud presentations (todas as figuras);
- ➔ Introdutory talk on RAMCloud.


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
bedankt
dziękuję
go raibh maith agat
sukriya
kop khun krap
terima kasih
grazie
arigatô
dakujem
takk
mochchakkeram
tapadh leat
спасибо
Баярлалаа
рахмат
dank je
misaotra
matondo
paldies
grazzi
mahalo
хвала
asante
manana
obrigada
tenki
murakoze
chokrane
mamnun
djiere dieuf
tau
дякую
taiku
sulpay
grazie
tanemirt
rahmet
diolch
dhanyavadagalu
shukriya
merce
мерси
trugarez
xiexie
ευχαριστώ
감사합니다
ありがとう
najis tuke
kam sah hamnida
rahmat
didi madloba
sagolun
mèsi
dėkuji
sobodi
enkosi
gratias ago
gràcies
chnorakaloutioun
vinaka
спасибі
blagodaram
mersi
kia ora
barka
welalin
tack
spas
taafetai lava
kiitos
dankie
dhanyavad
nanni
nandri
hvala
maururu
köszönöm
gracie
bayarlalaa
dankon
akun
aciu