

INSTITUTO SUPERIOR TÉCNICO

Curso em Engenharia Informática e de Computadores -

ALUNO

Número: 72531

Nome: Álvaro García Recuero

ORIENTAÇÃO

Nome: Luís Manuel Antunes Veiga

Categoria: Professor Auxiliar

Afiliação: Departamento de Engenharia Informática (DEI)

Nome:

Categoria:

Afiliação:

DISSERTAÇÃO

Título: HBase-QoD: Vector-Field Consistency for Replicated Cloud Storage

Data de provas: 20/09/2013

Idioma: Português

PALAVRAS CHAVES

Geo-Replicação, Bases de Dados NoSQL, HBase, Consistência Adaptável, Qualidade de Dados, Divergência de Réplicas

KEYWORDS

Geo-Replication, NoSQL Databases, HBase, Quality of Data, Tunable Consistency, Divergence-bounding

RESUMO (250 palavras)

Muitas das aplicações actualmente disponibilizadas em ambientes de computação em nuvem fazem uso de sistemas de armazenamento associativo chave-valor, tais como o BigTable, Cassandra, e muitos outros baseados em abordagens no-SQL para contornar as limitações de escalabilidade das bases de dados relacionais.

Para melhorar o desempenho e a escalabilidade, os sistemas de armazenamento são particionados por vários servidores, e centros de dados, com os dados de cada servidor replicados para assegurar disponibilidade. Além disso, parcelas do repositório chave-valor devem ser mantidas geo-cached tão perto quanto possível da periferia da rede, para maior desempenho e para reduzir a carga nos mega centros de dados.

Este trabalho tem como objectivo estender o HBase com políticas de caching e de replicação centradas no cliente, com um modelo de consistência baseado em limitação da divergência dos dados e na semântica das aplicações, que definimos como Quality-of-Data (QoD). Assim, os dados armazenados no HBase-QoD serão mantidos na réplica principal de um centro de dados com possivelmente várias réplicas secundárias denominadas region servers.

Globalmente, os dados podem obedecer a diferentes garantias de consistência e requisitos de sincronização, que serão aplicados na replicação entre centros de dados. Isto reduz o número de mensagens e largura de banda necessárias às réplicas para notificar aplicações de modificações nos dados ou actualizações. Isto, enquanto sendo capaz de fazer cumprir a semântica definida pelas aplicações de acordo com um modelo vectorial de consistência

ABSTRACT (250 words)

Many of today's applications deployed in cloud computing environments make use of key-value storage such as BigTable, Cassandra, and many other no-SQL approaches to overcome scalability limits of relational databases. Relevant open-source solutions include Apache HBase. Several works such as Percolator notify applications whenever data is updated by others (e.g., in the context of updating Google's web index).

For increased performance and scalability, such storage is partitioned across machines and data centers, and each node's data is replicated for availability therefore. Furthermore, fragments of the key-value store should be geo-cached as close as possible to the edge of the network location for increased performance and to reduce the load on mega data centers.

This work aims at extending HBase with client-centric caching and replication policies in regards to a consistency model based on data divergence bounds and user-defined application semantics, which we define as Quality-of-Data (QoD). Thus, data stored at HBase-QoD will be kept in the master of a data center with possibly several cached replicas on the slaves region servers.

Overall, the data may have different consistency guarantees and synchronization requirements that will be applicable to inter-replication with other master servers or clusters. This reduces the number of messages and bandwidth needed by master servers to notify applications of data changes and replica updates, while still being able to fulfill those data-defined semantics

JÚRI

Presidente:

Nome: Luís Eduardo Teixeira Rodrigues

Categoria: Professor Catedrático

Afiliação: Departamento de Engenharia Informática (DEI)

Vogais:

Nome: Johan Montelius

Categoria: Professor Associado

Afiliação: KTH - Royal Institute of Technology

Nome:

Categoria:

Afiliação:

Nome:

Categoria:

Afiliação: