
NAS 스토리지 기반의 데이터 분산처리 시스템

알고리즘에 관한 연구

장재명* · 강희범* · 정낙주* · 정희경*

*배재대학교 컴퓨터공학과

A Study on the NAS Storage-based Data Distributed Processing System Algorithm

Jae-Myung Jang* · Hee-beom Kang* · Nahk-ju Jeong* · Hoe-kyung Jung*

*Paichai University

E-mail : jaemyung0329@google.com, rkdgml90@naver.com, njjeong@gmail.com, hkjung@pcu.ac.kr

요 약

스토리지의 발달로 자동차 항공분야 등 실생활 어디에서든 스토리지가 활발히 활용되고 있다. 최근 빅데이터가 대두됨에 따라 많은 데이터를 스토리지에 저장하고 데이터를 처리하는 데이터 분산처리 연구가 활발히 진행되고 있다. 하지만 많은 사람들이 데이터를 동시에 요청할 때 병목현상이나 처리 속도가 느려지는 문제가 발생한다.

본 논문에서는 많은 데이터를 저장하고 처리해야하는 빅 데이터 분야에 사용될 것을 고려하여, 데이터 요청 시 보다 효율적으로 데이터를 처리하고 많은 데이터를 효율적인 관리가 가능한 데이터 경량화 처리 시스템 알고리즘을 제안한다.

ABSTRACT

Real life has been actively utilizing storage from anywhere automobiles and Aviation field etc to the development of storage. Recent Big Data is stored as a number of data storage and data distribution processing emerged in research has been actively conducted to process the data. But many bottlenecks or processing speed slows down when you request the data at the same time a problem arises.

In this paper consider should be used in the field of big data storing and processing large amounts of data, the process data more efficiently when the number of data request and data suggest that the weight-effective, manageable data processing system algorithm.

키워드

NAS, 빅데이터, 분산처리, Storage

1. 서 론

최근 스토리지의 발달로 자동차, 항공분야 등 실생활 어디에서든 스토리지가 활발히 활용되고 있는데, 최근 빅 데이터가 대두됨에 따라 많은 데이터를 스토리지에서 실시간으로 처리되어 데이터들의 크기는 점점 커져가고 처리속도는 늘어나고 있다.

이에 따른 문제점이 발생하는데 클라우드 스토

리지에서 사용자들이 특정 데이터 호출이 많아질 경우 데이터 병목현상의 발생을 야기 하고 스토리지에 부하가 자주 일어난다[1,2].

이런 문제를 해결하기 위해 사용자의 요청을 분산하고 처리하여 저비용/고성능의 스토리지 개발이 진행 중이다. 방법들 중 NAS (Network Attached Storage)가 각광 받고 있고 데이터 중복 제거, 압축 등 데이터 안정화 기술의 연구가 활발히 진행되고 있다.

본 논문에서는 많은 데이터를 저장하고 처리해야 하는 빅 데이터 분야에 사용될 것을 고려하여, 데이터 저장 시 보다 효율적으로 데이터를 처리하고 많은 데이터가 스토리지에 저장되더라도 효율적인 데이터 관리가 가능해지는 데이터 경량화 처리 시스템 알고리즘을 제안한다.

II. 관련연구

2.1 NAS의 연결방식

스토리지 구성 기술 측면에서 볼 때 가장 안정적으로 자리 잡은 기술은 NAS(Network Attached Storage)이다. 저장장치들이 서버에 직접 연결되어 액세스 되던 기존의 형태에서 벗어나 서버와 저장장치들을 별개의 전용 네트워크로 연결함으로써 여러 대의 서버가 저장장치를 공유할 수 있을 뿐만 아니라 전용 네트워크를 이용하기 때문에 기존의 직접 연결방식(Direct Attached storage)에 비해 안정적으로 빠른 액세스가 가능해 졌다. NAS는 네트워크로 연결되어 데이터를 언제든지 사용할 수 있다는 장점을 가지고 있지만 사용자가 데이터를 동시에 처리할 때는 느려지거나 접속이 원활하지 않는 단점을 가지고 있다[3,4].

III. 시스템 설계

기존 클라우드 스토리지 아키텍처는 동시에 많은 클라이언트가 데이터 Request를 요청하게 되면 스토리지 서버에서는 병목현상 및 부하가 일어난다.

본 논문에서는 NAS라는 스토리지 시스템 구조를 기본으로 구성 하였으며 자주 쓰는 데이터들을 HOT Data로 구분하여 따로 저장한 뒤 데이터를 분산처리 한다. 그림1 과 같이 데이터를 분산처리 하는 데이터 처리 구성도이다.

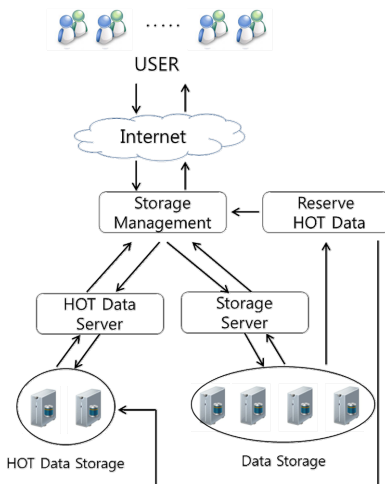


그림 1. 클라우드 스토리지 데이터 전송 구성도

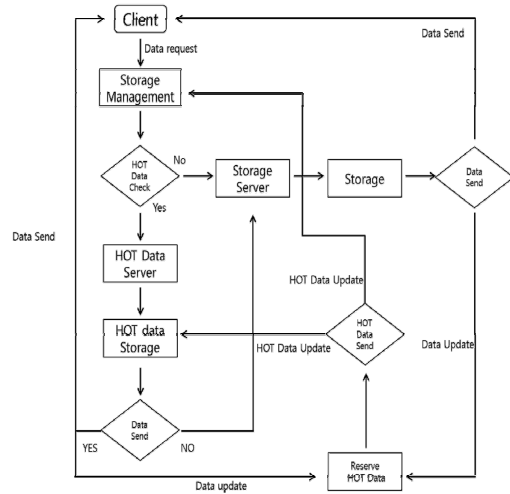


그림 2. 클라우드 스토리지 알고리즘 순서도

본 논문에서 제시하는 클라우드 스토리지 알고리즘 순서도는 그림 2. 과 같이 먼저 클라이언트가 서버에 데이터를 Request 하면 Srotage Management에서 데이터가 HOTS Date Storage에 저장되어 있는지 여부를 확인한 뒤 전송을 요청 한다. HOTS Date Storage에서는 클라이언트로 데이터를 보내주어 분산 처리하여 데이터가 집중되는 현상을 미연에 방지할 수 있어 데이터 병목현상을 최소한으로 줄일 수 있다. 또한 HOTS Data Storage안에 데이터가 없다면 다시 Storage Server로 요청을 보내어 다시 Storage에서 데이터를 보낼 수 있도록 해야한다. 클라이언트로 데이터를 보냄과 동시에 Reserve HOTS Date로 Count 값을 1을 보내주고 값을 저장 하게 되는데, 만약 그 값이 임계값이 되면 그 데이터는 Storage Management페이지에 값의 주소, HOTS Data Storage에는 스토리지의 데이터가 복제가 되어 실시간으로 데이터가 HOTS Data Storage에 저장이 된다.

IV. 결론

최근 클라우드 시스템에 빅데이터가 접목되어 많은 데이터들이 스토리지에 저장이 되면서 이에 관련된 연구들이 활발히 이루어지고 있으며 기존에 DAS, NAS, SAN의 기술들 및 2가지의 장점을 합치는 기술들이 부각 되고 있다. 하지만 많은 클라이언트들이

데이터를 요청하면 병목현상이 일어나 처리속도가 느려지는 단점이 발생한다.

본 논문에서는 많은 데이터를 저장하고 처리해야 하는 빅 데이터 분야에 사용될 것을 고려해서, 데이터를 Storage Management로 관리하여 데이터 병목현상을 미연에 방지 한다. 결과적으로 데

이더 처리의 효율성을 증대 시키는 효과를 기대할 수 있다.

향후 연구로는 제시한 알고리즘을 사용 및 NAS를 융합하여 데이터 처리 알고리즘 구현 및 효율성을 검증해야 한다.

참고 문헌

- [1] <http://crisco.tistory.com/292>
- [2] <http://skymeet.tistory.com/32>
- [3] J. S. Lee and Y. K. Yen and H. C. Roh and Y. M. Yoon and S. H. Park “A Survey of Sequence Alignment Algorithms with Distributed System” in Journal of KIIT. Vol. 12, No. 7, 2014
- [4] S. I. Moon and J. K. Kim and T. Y. Kim and J. S. Choi and J. S. Lee “Job Classifying method based on Data Traits for Increased Effic” in Journal of the Korea Society for Simulation Vol. 23, No. 4, pp. 219-228, 2014