

Science DMZ를 활용한 빅데이터의 효율적 공유 방법 연구

최원준, 김재훈, 김순영, 김정환
한국과학기술정보연구원, 정보기반실

e-mail : cwj@kisti.re.kr, jay.kim@kisti.re.kr, maya@kisti.re.kr, kimjh@kisti.re.kr

A Study of Big Data Transmission using Science DMZ

Won-Jun Choi, Jae-Hun Kim, Sun-Young Kim, Jeong-Hwan Kim
Korea Institute of Science & Technology Information

요 약

제 4 차 산업혁명(클라우드, 빅데이터, Internet of Things, 인공지능 등)과 관련된 다양한 기술이 개발되고 연구가 되면서 네트워크를 사용하는 다양한 기기(모바일, 클라우드 서버 등)에서 축적되는 데이터의 양도 점점 증가하고 있다. 데이터 전송 서비스를 지원하는 기관이나 데이터 전송을 위한 인프라를 지원하는 기관에서는 종종 데이터가 빅데이터가 되어갈수록 기관 간의 효율적인 데이터 전송이 어려워 직접 방문하여 필요한 정보만 전달 받기도 한다. 초고속 네트워크 시대에 우리는 과학기술정보의 효율적인 공유를 위한 신속한 전송이 필요하게 되었고 이러한 방법을 Science DMZ 환경에서 문제를 풀어 보고자 한다.

1. 서론

최초의 산업 혁명은 유럽과 미국에서 18 세기에서 19 세기에 걸쳐 일어났다. 주로 동력 발생원으로 증기기관을 사용하여 생산성을 향상시켰으며, 제 2 차 산업혁명은 동력 발생원으로 모터를 사용하여 작게 만들 수 있고, 쉽게 제어 가능해서 생산성 향상시켰으며, 제 3 차 산업혁명은 전자회로, 정밀제어에 의한 생산성 향상을 꾀하였다. 최근에 이슈로 떠오른 제 4 차 산업혁명은 초연결성, 초지능성에 의한 생산성 향상으로 인공지능, 로봇공학, 사물 인터넷, 무인 운송 수단, 3D printing, 나노 기술, 연결 및 표시 기술 등이 있다. 이러한 기술들이 발달함에 따라 데이터의 전송을 필요로 하는 기기(모바일, 클라우드 인프라 등)에서는 더욱 더 많은 데이터의 전송 요구가 증가되었고 이러한 데이터의 인프라를 지원하는 기관에서는 기관 간의 데이터의 효율적 공유 방법이 필요하게 되었다. 현재 천문 데이터와 같은 경우에는 빅데이터가 많기 때문에 직접 방문해서 필요한 데이터를 담아 가지고 오는 형편이다. 또한, 보안의 이슈로 트래픽이 방화벽을 거치게 되면 속도의 delay가 발생하여 데이터 전송의 효율성에 문제가 발생한다. 이와 같은 문제를 해소하기 위해서 방화벽의 물리적 트래픽 제어로 전송에 제약을 덜 받는 Science DMZ기술이 등장하게 되었고 이를 토대로 데이터를 가지고 있는 각 기관에서 데이터를 직접 공유하게 되면 직접 방문하지 않아도 적은 시간에 효율적으로 데이터를 전송할 수 있다. 미국에서는 ESNET을 중심으로 미국 서부지역 유수의 대학과 연구소들에 대한 Science DMZ화를 통하여 빅데이터, 슈퍼컴퓨팅, 초고성능

전송 기반의 첨단 응용 연구들을 선정하여 효과적인 데이터 집중형 과학 응용연구를 활성화하고 있다. 본 연구에서는 직접 Science DMZ에서 사용되는 데이터 전송 노드(DTN)를 구현해 보고 실제 전송 속도를 측정하여 데이터 전송의 효율성을 가늠해 보고자 한다.

2. 과학기술정보 빅데이터의 증가

국내뿐만 아니라 해외에서도 학술정보 유통을 위한 과학기술정보 데이터 서비스 사업에 대한 지원이 확대되고 있다. 과학기술정보를 활용하여 서비스되고 있는 플랫폼은 다음과 같다. 국가과학기술정보센터(NDSL)[1]은 NDSL은 산학연 연구자들에게 고품질 정보를 제공하여 국가 과학기술 혁신에 기여하고자 개발된 과학기술정보 서비스 플랫폼으로, 논문·특허·보고서·동향·저널/프로시딩·연구자·연구기관 등 약 1 억건 이상의 콘텐츠에 대한 검색 및 콘텐츠 유형별 전문검색 서비스를 제공하고 있다. 학술정보통합관리시스템(KSCD)[2]은 KISTI 과학기술정보센터에서 구축 및 운영하는 국내 과학기술분야 학술콘텐츠 데이터베이스이다. 전자정보국가컨소시엄(KESLI)[3]은 국내외 전자정보 공급사들로부터 다양한 정보를 확보하고 국내 학연 산 기관들이 보유한 정보를 공동 활용할 수 있는 체계를 구축 및 운영함으로써 국내 연구자들이 필요로 하는 다양한 정보를 신속하게 제공할 수 있는 지식정보서비스를 제공하고 있다. 또한, 과학기술정보 데이터에는 문헌정보도 있지만 과학 실험데이터, 소스데이터도 존재한다. 천문분야의 실험데이터를 예를 들면 그 규모가 방대해서 일반 TCP로 구성된 랜 환경에서의 데이터 공유는 데이터 전송 효율성에

제약이 따른다. 이와 같이 과학기술정보 데이터가 늘어갈수록 개인 간, 기관 간 또는 국가 간에 빅데이터를 공유하는 요구는 증가하고 있다.

2.1 클라우드 환경의 데이터 공유 시스템

연구자 간의 또는 기관 간에 과학기술 빅데이터를 공유하고자 하는 움직임은 급속도로 빨라지고 있다. 모든 종류의 과학적 지식을 개방하고 공유하는 오픈 사이언스라는 개념이 도입되었고 d-collection, d-space 등의 학술정보 공유서비스가 지원되고 있다. 하지만, 실질적인 실험데이터, 클라우드 스토리지 등의 정보에 대한 공유에는 기관 간의 정책문제, 방화벽 문제 등으로 인해서 전송 지연이 발생한다. 본 논문에서는 클라우드 환경에서 데이터를 효율적으로 공유하기 위해 방화벽이나 데이터 교환 정책을 일일이 거치지 않고 바로 전송할 수 있는 Science DMZ 네트워크의 DTN 을 사용하여 효율성을 검증하였다.

2.2 성능 평가

실험은 다음과 같이 진행하였다. 서울과 대전에 DTN 서버를 셋팅하였고 10Gbps 로 네트워크를 튜닝하였다. 성능 검증은 PerfSONAR 의 bwctl 도구를 사용하였다. 일반 네트워크로 10G 과학기술정보 데이터를 대전에서 서울로 전송하였을 때 속도는 100Mbps 정도가 최대가 되었으며 DTN 노드로 전송하였을 경우에는 9.31 Gbits/sec 정도의 성능을 볼 수 있었다. 튜닝은 Esnet 의 튜닝 가이드라인을 참조하였다.

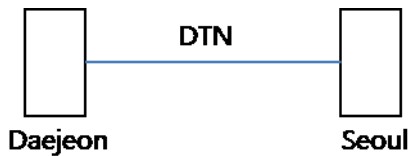


그림. 1 DTN 네트워크 구성도

실험을 위해 그림. 1 과 같이 DTN 노드를 튜닝하였다. 파일 전송 크기를 10G 로 설정하였을 경우 iperf3 의 성능 평균 속도는 9.31 Gbits/sec 로 나왔고 ftp 로는 초당 117MB, Gridftp 로는 164.4 MB/s 의 성능이 나왔다. 방화벽이나 데이터 교환 정책을 사용하지 않고 신뢰적인 네트워크를 구성한다면 충분히 높은 성능의 네트워크 속도를 가져올 수 있을 것으로 생각되며 기관 간의 과학기술정보의 공유 속도가 증가하면 클라우드 환경에서의 데이터 공유가 점점 더 활발하게 진행될 것이라 생각된다.

2.3 결론

과학기술이 발달함에 따른 데이터의 증가는 필수가 되었고 이러한 데이터의 효과적인 공유문제 또한 필수가 되었다. 다양한 분야에서 생성된 빅데이터를 기관 간에 효과적으로 공유하기 위해서는 Science

DMZ 네트워크에서의 전송이 필요해 보인다. 초연결 시대를 살아가는 현 시점에서 사람과 사람, 기관과 기관, 장치와 장치 등의 다양한 플랫폼에서의 신뢰적이고 보안이 강화된 네트워크는 필수로 요구되고 있다. 오픈 사이언스의 개념을 바탕으로 이러한 기술이 앞으로 기관 간의 과학기술정보를 좀 더 효율적으로 공유할 수 있는 기반이 될 것으로 생각된다.

참고문헌

- [1] 강남규, 신용주, 박근철, 주원균 외(2010), NTIS-NDSL 활용한 국가R&D논문 수집 방안에 관한 연구, 한국인터넷정보학회 학술발표대회 논문집, 2010.10, 231-232.
- [2] 김병규, 강무영, 최선희 (2006), 협회 기술정보 관리 및 유통 시스템 구축에 관한 연구. 정보관리연구, 37(3), 117-137.
- [3] 전자정보국가컨소시엄, Kesli.or.kr
- [4] 박형우, 윤희준, 장해진(2012), The Survey on Science DMZ for Big Science Data Transfer, 한국콘텐츠학회 종합학술대회 논문집, 403-404.
- [5] 문정훈, 이민선(2016), 데이터 집중형(Data-Intensive Science) 과학을 위한 ScienceDMZ 기반 빅데이터 전송 연구, 한국정보과학회 학술발표논문집, 1359-1361.