

# 과학기술 데이터댐 네트워크 서비스 모델 연구

노민기

한국과학기술정보연구원

mknoh@kisti.re.kr

## A Study on Network Service Model for S&T Data Dam

Minki Noh

Korea Institute Science Technology and Information

### 요약

과학기술 데이터댐은 대용량 데이터를 중심으로 다수의 연구 커뮤니티에서 관련 연구를 수행할 수 있도록 대용량 데이터의 수집과 분석, 서비스를 제공한다. 이와 같은 일련의 데이터 서비스를 통해 연구자들은 자신의 연구 현장에서 데이터의 특성과 정보를 획득하고, 이를 쉽고 효율적으로 활용하여 연구 생산성과 품질을 높인다. 연구자들이 대용량의 데이터를 획득하고자 할 때, 직접적이고, 유리한 방식은 데이터의 생산지에 자신의 시스템을 직접 연결하고, 실시간으로 데이터 생산지역의 시스템에 접근하여 해당 데이터를 수신받아 활용하는 방식이다. 하지만, 다수의 사용자와 공동으로 데이터를 활용할 경우 이와 같은 직접적인 방식이 어려우며, 또한 일반적인 업무에 함께 활용되는 네트워크로 연결된 대용량 데이터와 일반적 인터넷 데이터를 복합적으로 처리하는 일반적인 네트워크 구조상 효율적인 네트워크 환경을 구성하기는 어렵다. 본 논문에서는 이러한 일반적인 연구 환경을 대상으로 고성능의 송수신 네트워크 구조인 대용량 데이터 전송구조(ScienceDMZ)의 모델을 과학기술 데이터댐 서비스에 적용하여 과학기술 데이터댐의 네트워크 성능과 데이터서비스 향상과 확대를 목적으로 하고 있다.

### I. 서론

현재의 과학기술은 개인능력의 우수성을 통한 성과 창출보다는 동일한 목적의 연구주제를 중심으로 다수 혹은 복수지역의 연구자들과 함께 공동으로 수행하는 것으로 변하고 있다. 또한, 공동 연구 수행을 위한 데이터 수집, 계산자원, 교류와 공동 활용을 위한 플랫폼과 인프라의 발전이 함께 필요하다.

공동의 자원과 연구플랫폼의 발전은 데이터의 대용량화를 선도하였으며, 이는 데이터를 처리하는 계산자원과 저장하는 저장 자원뿐 아니라, 데이터를 송수신하는 네트워크의 고속화와 대역폭의 확대로 함께 발전하고 있다. 또한, 다수의 협력으로 수행되는 공동연구는 생산→처리/가공→활용의 일련의 과정을 분류하여, 이 과정에서 필요로 하는 핵심적인 기술과 인프라, 플랫폼을 공동의 연구 환경으로 설정하여 역할별로 생산자, 소비자뿐 아니라, 가공과 유통, 전달의 단계의 중요성은 연구환경의 전체적 품질의 고도화와 함께 매우 중요한 연구 환경으로 인식되고 있다.

더 나아가 이러한 생태계를 중심으로 데이터를 활용하는 연구자(기관, 커뮤니티)들에게 자신의 연구활동에 사용될 데이터의 위치뿐 아니라, 위치한 데이터 중 가장 연구자들이 접근, 활용성이 우수한 정보를 제공하는 것은 과학기술 연구 생산성 향상과 향상된 결과 취득에 많은 영향을 미치고 있다.

대용량의 데이터를 연구에 활용할 때 가장 직접적이고, 유리한 네트워크 방식은 데이터의 생산지와 소비자를 직접 연결하고, 연구자가 데이터가 필요할 때, 실시간으로 생산지에 접근하여 획득하는 방식일 것이다. 하지만, 일반적인 연구기관에서는 현재 대용량의 데이터를 수신하여 연구 활동을 수행하는 것뿐 아니라, 일반적인 업무 즉, 이메일, 전자적 시스템, 조직과 재정정보, 연구정보, 개인정보등을 처리하는 복합적으로 처리하는 네트워크에 연결되어 있다. 이러한 네트워크는 연구기관에 속한 일부 대용량 데이터 중심의 연구자가 안정적이고 고성능으로 과학기술데이터댐의 서

비스를 충분히 활용하기 어려운 환경이다.

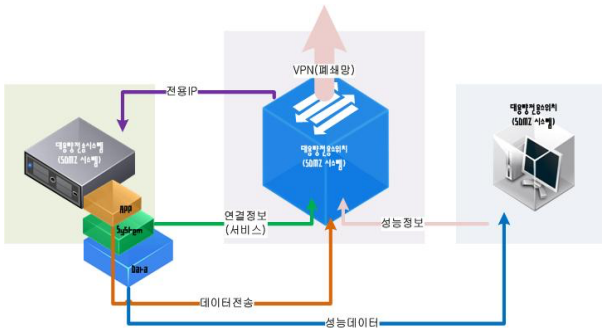
본 논문에서는 이러한 문제를 인식하여 제안된 대용량데이터 전송구조(ScienceDMZ)의 우수한 네트워크 구조의 장점을 과학기술데이터댐의 송수신 단계와 서비스 단계에 접목하여 대용량데이터를 활용하는 연구환경의 네트워크 환경과 연구수행의 효율성 향상 지원을 목적으로 하고 있다.

### II. 본론

일반적으로 네트워크는 많은 데이터를 분석하고 저장된 데이터를 공유하는 인프라로 활용되지만, 범용 업무망(General Purpose Network)은 대용량 데이터 전송에 최적화되지 않은 네트워크이기 때문에 데이터 중심형 연구의 요구사항을 지원할 성능을 가지고 있지는 않다. 과학자들이 범용 업무망에서 데이터 중심의 연구 활동을 수행했을 때 그 결과는 낮은 전송성을 경험하거나, 데이터를 수집하고 송수신하는데 장시간의 연구 시간을 소모하여야 한다.

이러한 문제를 해결하기 위해 미국 에너지성 연구망인 ESnet에서는 ScienceDMZ 전송모델을 통해 대용량 데이터 활용 기준의 네트워크 송수신 체계를 설계, 운영과 보안이라는 구성요소로 나누어 제안하였다. 전통적인 DMZ는 기관의 내부망과 보안이 필요한 일부 서비스 즉, 외부에 노출이 필요한 홈페이지, 이메일, DNS 서버 등과 같은 서버들에 따로 네트워크로 구성하여 제공하는 구간설정을 말한다. ScienceDMZ에서는 대용량 데이터 중심의 협업 형태의 과학 응용연구를 지원하기 위해 고성능 네트워크 구간을 설계하여 연동하고, 연구자들이 접속하여 대용량 데이터에 접근, 전송할 수 있는 DTN(Data Transfer Node)으로 구성 설계된다.

이러한 Science DMZ 전송모델은 송신지역과 수신지역간의 대용량 데이터의 전송을 원활하고 안정적으로 보장할 뿐 아니라, 송수신 시스템에서 전송량과 성능의 정보, 시스템 정보 등을 획득하여 실질적으로 대용량 데이터의 획득과 전송처리를 빠르고 안정적으로 처리한다.



<그림 1> 대용량데이터전송구조(ScienceDMZ)

과학기술 데이터댐의 구성은 이러한 ScienceDMZ의 구조를 활용하여 대용량 데이터를 연구에 활용할 수 있도록 서비스 구조, 수집-저장/분석-제공의 단계마다 네트워크 구조와 DTN을 설계하여, 다수의 연구자들이 데이터댐에 접속하여 데이터의 형태와 특성이 분류된 서비스를 제공받고, 연구에 필요한 데이터를 연구자 최종단계에서 빠르게 활용할 수 있도록 한다. 이를 위해, 과학기술데이터댐의 서비스 구조인 생산→처리/가공→유통→활용(제공)과 ScienceDMZ 전송구조를 네트워크의 특징에 따라 정리하면 다음과 같다.

- ① 대용량데이터의 수집시 성능과 품질이 일반망보다 우수하고, 연구시간 혹은 대용량데이터를 연구에 활용하는 사용시간 내에 안정적인 망이 필요
- ② 대용량데이터 제공시 범용적이고 다수 시간에 네트워크 활용보다는 연구데이터의 크기와 송수신의 목적, 데이터의 위치에 따른 특정 데이터에 따른 네트워크 환경이 필요
- ③ 대용량데이터의 송수신 환경에서 특정한 연구목적을 충족하는 특수한 망 기술의 적용과 보안이 필요

과학기술 데이터댐을 구성하는 송수신과 저장의 역할을 담당하는 시스템은 DTN과 연결하여 데이터를 발생하는 지점에서 안정적이고 빠르게 데이터를 수신하여 저장한다. 저장된 데이터를 제공하는 구간에서는 직간접적으로 연구자들의 요구, 대역폭, 흐름, 전송 기술을 분석하여 의미 있는 정보로 제공하는 것이 필요하다. 이를 위해, 과학기술 데이터댐을 데이터 생산(Data Collector)지역, 데이터 집결지/저장지(Central Data repository)와 데이터 소비지역으로 나누었을 때 각 구간의 요구는 다음 3가지의 구조라고 요약할 수 있다.

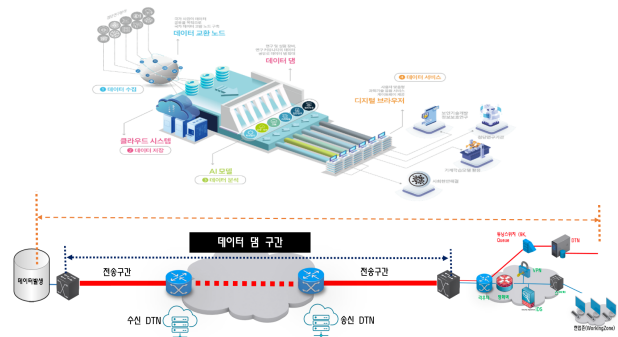
- ① 1단계 : 네트워크 연결 (Network Connection)
  - 대용량 네트워크로 접속가능한 기가급 액세스망
  - 지역/물리적 위치 중심의 네트워크 연결, IP 연결
- ② 2단계 : 논리적 연동 (Logical Network Peering)
  - 라우팅과 별개의 논리망으로 일정한 대역폭과 변동가능한 대역폭 구간
  - 패스(Path)구성 및 터널링 구성.네트워크 구조 트래픽 흐름 제어
- ③ 3단계 : 어플리케이션 중심 네트워크(Applications Network)
  - 사용자들에게 데이터 정보가 제공
  - 실시간 데이터정보, 송수신대상, Path, 대역폭, 흐름정보 제공
  - ①-③ 구간의 성능 및 보안모니터링 (관제 및 정보제공)

위의 구간에서 2단계 논리적 연동구간에 Science DMZ의 전용구간모델, 3단계와 성능 및 보안모니터링 구간에 DTN 모델을 적용하면 아래 <표 1>와 같다.

<표 1> 전송모델 : 데이터발생지역 - 데이터댐(수신 DTN)-데이터댐(송신 DTN)-데이터활용지역

	전송구간		전송구간	역할
SDMZ (1)	1단계	↔	2단계	데이터수집
DTN(1)	1단계	↔	2단계	데이터수집/저장
SDMZ (2)	3단계	↔	3단계	데이터제공
DTN(2)	3단계	↔	3단계	데이터 제공
DTN(1),(2)	1단계	↔	3단계	정보제공

위의 과학기술 데이터댐의 네트워크 구간 (물리적단계-논리적단계-어플리케이션구간)을 중심으로 구성된 네트워크 구간 중 논리망 연동구간과 어플리케이션 중심 네트워크 구간에 SDMZ모델의 DTN을 구성하여 각각 데이터 입력과 출력의 게이트웨이 역할을 수행하며, 마지막으로 성능관측 구간에서는 SDMZ의 성능측정과 정보를 제공하는 PerfrSONAR를 통해 각 네트워크의 성능과 품질을 모니터링 하는 역할을 수행하도록 설계한다.



<그림 2> 과학기술데이터댐 ScienceDMZ 적용모델

위 <그림 2>와 같이 과학기술 데이터댐과 ScienceDMZ의 전용기술과 DTN을 적용하여 네트워크 구조를 설계하면, 네트워크 환경분석 요소를 중심으로 최적화된 네트워크 구성이 가능할 뿐 아니라, 통합적인 데이터 서비스를 제공할 수 있는 네트워크 전송모델 구성이 가능하다.

### III. 결론

본 논문에서는 과학기술 데이터댐의 생산→처리/가공→유통→활용(제공)의 데이터 서비스 과정과 데이터를 송수신하는 과정을 ScienceDMZ의 전용기술을 통해 보강하여 데이터중심의 연구환경을 제공하는 과학기술 데이터댐의 역할을 강화하고 서비스를 확대를 목적으로 하고 있다. 공통의 연구환경에서 연구자들의 연구수행 지원은 하나의 연구환경을 중심으로 공동 연구생산의 우수한 결과를 취득하려는 생태계를 형성할 수 있기 때문이다.

과학기술 데이터댐은 다수의 연구 커뮤니티에서 데이터중심형 연구를 수행할 수 있도록 대용량 데이터와 함께 이를 연구자들이 쉽게 획득 할 수 있도록 분석, 정보 서비스를 제공한다. 이와 같은 데이터 서비스를 통해 연구자들은 자신의 연구 현장에서 데이터의 특성과 정보를 획득하고, 이를 쉽고 효율적으로 활용하여 연구 생산성과 품질 향상을 기대할 수 있다. 현재 다수의 연구자들은 공동으로 일반적인 업무에 함께 활용되는 일반적인 네트워크로 과학기술 데이터댐을 활용한 대용량 데이터의 획득과 일반적 인터넷 사용을 동시에 처리하는 어렵다. 이러한 환경을 대상으로 고성능의 송수신 네트워크 구조인 대용량 데이터 전송구조(ScienceDMZ)의 모델을 네트워크와 전송부분에 적용하면 과학기술데이터댐의 네트워

크 성능과 데이터서비스가 효율화 될 것 이다. 또한, 데이터의 송수신과 연구자를 연결하는 글로벌 수준의 네트워크 인프라는 협업 연구공간에서 연구자, 생산자, 소비자가 공존하고, 각 연구수행에 필요한 연구자원, 계산 자원, 네트워크자원이 결합, 융합 등이 유리한 환경이 될것으로 기대된다.

고에너지물리, 천문우주, 핵융합, 기상기후 등 데이터중심의 연구는 보장된 고속의 전송 환경을 요구하고 있으며, 이는 다수의 기관 또는 연구자들이 공통으로 충족되어야 참여하는 협업연구의 품질과 성과가 보장된다. 이를 위해 과학기술 데이터댐의 주요 서비스 단계에 ScienceDMZ의 네트워크 구조를 재구성하여 모델화하여 제공된다면, 과학기술 데이터댐의 송수신 단계에서의 네트워크 성능뿐 아니라, DTN을 데이터 서비스 단계에 제공하여 데이터의 정보획득과 수신 성능의 향상과 과학기술 데이터댐의 서비스 대상 확대를 기대할 수 있다.

## 참 고 문 헌

- [1] fasterdata.es.net.com, last modified Sept. 18, 2019, accessed Sept. 18, 2019, <https://fasterdata.es.net/science-dmz/>
- [2] KREONET 첨단연구사례집(2005-2020)
- [3] globus.org, last modified Sept. 18, 2019, accessed Sept. 18, 2019. <https://www.globus.org/>
- [4] perfsonar.net, last modified Sept. 12, 2019, accessed Sept. 18, 2019, <https://www.perfsonar.net/>