iRODS를 이용한 대용량 전자기록물 관리 시스템

(Massive Electronic Record Management System using iRODS)

한 용 구 † 김 진 승 † 이 승 현 † 이 영 구 ††
(Yongkoo Han) (Jinseung Kim) (Seunghyun Lee) (Young-Koo Lee)

요 약 정보통신의 발전으로 인한 전자 기록물의 등장은 기록물 관리 체계에 많은 변화를 가져 왔다. 전자 기록물의 등장으로 인한 기록물관리의 가장 큰 변화는 종이기록물 등의 물리적인 기록물에 대한 수동적인 관리체제에서 컴퓨터 시스템을 통한 전자형식 기록물의 자동화된 기록 관리체제로의 전환과 대량·대용량으로 발생하는 전자 기록물에 대한 효율적인 관리이다. 데이터 그리드 소프트웨어인 iRODS (Integrated Rule-Oriented Data System)는 규칙 기반 데이터 그리드 시스템으로서 지리적으로 분산된 컴퓨팅 자원을 네트워크로 연동하여 가상화를 통해 대용량 저장 인프라를 구성한다. 특히, 이 시스템은 규칙을 정의함으로써, 입력되는 전자 기록물에 대한 데이터를 분산 및 백업하는 등의 조작이 가능하다. 따라서 iRODS는 대량, 대용량으로 발생하는 전자기록물을 자동적으로 관리하는 전자기록물 관리시스템의 구축에 적합한 기술이다. 본 논문은 규칙 기반의 그리드 시스템인 iRODS를 이용하여 전자 기록물 관리 시스템을 설계 및 구현하였다. 본 논문에서 구축한 전자 기록물 관리 시스템은 iRODS의 규칙을 이용하여 기록물 유형에 따라 데이터를 자동 분산하며 데이터를 자동으로 백업한다. 그리고 iRODS의 메타데이터를 관리하는 iCAT DB를 이용하여 전자 기록물의 메타 데이터 저장 및 검색 기능을 제공한다.

키워드 : 기록물관리 시스템, OAIS 참조모형, 데이터 그리드, iRODS

Abstract The advancement of electronic records brought great changes of the records management system. One of the biggest changes is the transition from passive to automatic management system, which manages massive records more efficiently. The integrated Rule-Oriented Data System (iRODS) is a rule-oriented grid system S/W which provides an infrastructure for building massive archive through virtualization. It also allows to define rules for data distribution and back-up. Therefore, iRODS is an ideal tool to build an electronic record management system that manages electronic records automatically. In this paper we describe the issues related to design and implementation of the electronic record management system using iRODS. We also propose a system that serves automatic processing of distribution and back-up of records according to their types by defining iRODS rules. It also provides functions to store and retrieve metadata using iRODS Catalog (iCAT) Database.

Key words : Electronic Record management System, OAIS reference model, Data grid, iRODS

· 본 연구는 행정안전부 국가기록원의 지원을 받아 기록물 보존기술 연구개발 (R&D) 사업의 일환으로 이루어졌으며, 이에 감사드린다.

† 학생회원 : 경희대학교 컴퓨터공학과
ykhan@khu.ac.kr
jskim@cctank.com
niceshut@khu.ac.kr
†† 종신회원 : 경희대학교 컴퓨터공학과 교수
yklee@khu.ac.kr

논문접수 : 2009년 12월 29일
심사완료 : 2010년 5월 25일

정보과학회논문지: 컴퓨팅의 실제 및 레터 제16권 제8호(2010.8)

1. 서 론

정보통신의 발전으로 인한 전자 기록물의 등장은 기록물 관리 체계에 많은 변화를 가져 왔다[1]. 현재 여러 국가들이 전자기록물 관리체계를 도입하여 전자정부를 구축하였으며, 그 과정에서 컴퓨터와 네트워크 등의 정보통신기술을 이용하여 기관의 업무를 보다 효율적으로 처리 할 수 있는 정보화된 사무환경으로 변화시키는 계기가 되었다. 이러한 사무환경의 변화는 기록물의 보관을 위한 물리적 위치와 보관을 위한 방법의 변화를 야기하였고, 그에 따른 정보화된 사무환경에 적합한 기록물 관리 체계를 요구하게 되었다[2]. 전자형식 기록물은 정보통신기술의 발달에 따라, 생산되는 기록물이 대량으로 발생되고 저장된다. 또한 각각의 기록물들은 문서기반의 기록물 위주에서 멀티미디어 데이터로 그 범위가 확장되어 개별적인 기록물의 크기가 대용량으로 변화하고 있으며, 이러한 전자 기록물의 대량·대용량화에 따라 다음과 같은 관리의 문제점이 발생한다. 첫째, 전자 기록물은 분산된 하부 조직 또는 지역적으로 분산된 곳으로부터 발생하기 때문에 관리자들에 의하여 각기 다른 방식으로 관리되고 있다. 따라서 정형화된 관리 정책이 적용되지 않아 정책의 변화 또는 추가에 유연하게 대처하기가 어렵다. 또한 관리자에 의한 기록물 손실 또는 관리의 실수가 발생할 수 있다. 둘째, 하부 조직 및 지역적으로 업무상 특성 또는 전자 기록물의 형태에 따라 각기 관리되기 때문에 데이터 관리 정책 등의 통합적 관리가 어렵다. 또한 시스템의 저장 공간이 한계에 도달한 경우 저장시스템의 규모를 확장하기 위해서는 많은 노력과 예산이 필요하게 된다. 그리고 전자기록물들은 점차 대량·대용량의 성격으로 변화하고 있기 때문에, 저장시스템의 규모를 예측하기 어렵다. 따라서 이와 같은 문제점들을 해결할 수 있는 대량·대용량 전자 기록물에 대한 효율적인 관리체계가 요구된다.

첫 번째 문제를 해결하기 위한 방안으로는 대량의 전자기록물의 저장을 자동화하는 "프로세스 자동화"에 대한 연구인 워크플로우[3]와 업무 프로세스 관리(BPM)[4]가 있으며, 이를 이용한 자동화 시스템 구축 등을 고려할 수 있다. 그러나 이러한 연구들은 자동화 시스템 도입 부분에서 이기종 시스템과의 호환성에 취약하며 정책이나 상황별 유연성의 확보가 어려워 다양한 전자기록물 생성 경로와 관리 정책, 시스템적 차이 등을 극복하는데 한계가 있다. 두 번째 문제를 해결하기 위한 방안으로는 대량·대용량의 데이터를 논리적으로 통합된 분산된 공간에서 처리하기 위한 분산처리 기술이 있으며, 대표적으로 클라우드 컴퓨팅과 데이터 그리드를 통하여 구현할 수 있다. 하지만 클라우드 컴퓨팅은 개인 PC나 서버 등의 컴퓨터 연결을 통해 거대한 서버를 구축하여 단말기나 장소에 구애받지 않고 원하는 작업을 수행할 수 있도록 하는 컴퓨팅 기술로서 데이터의 보존 관리보다는 원격 응용 서비스에 적합한 분산 처리 기술이다. 반면에 그리드 컴퓨팅은 지역적으로 분산된 다수의 이기종 컴퓨팅 자원을 네트워크로 연동하여 가상의 대용량 고성능 컴퓨팅 시스템을 구성하는 분산 컴퓨팅 기술이기 때문에 전자기록물의 분산처리에 클라우드 컴퓨팅보다 적합하다. 이러한 이유로 데이터 그리드 컴퓨팅 기술은 분산된 컴퓨팅 자원을 데이터 공유와 저장을 목적으로 사용하기 때문에 데이터 공유, 디지털 라이브러리, 디지털 아카이브 구축 등에 활용되어 왔다[5]. 하지만 이러한 그리드 컴퓨팅 기술은 업무 프로세스의 자동화를 지원하는 기능이 미비하다.

이러한 자동화 기능과 분산 기술, 가상화 기술을 지원하는 미들웨어를 Data Intensive Cyber Environments 연구그룹에서 개발하였다. 데이터 그리드 소프트웨어인 iRODS(Integrated Rule-Oriented Data System)[6]는 규칙(rule) 기반 데이터 그리드 시스템으로서 지리적으로 분산된 컴퓨팅 자원을 네트워크로 연동하여 가상화를 통해 대용량 저장 인프라를 구성하는 기술이다. iRODS의 규칙을 이용하여 전자 기록물의 저장, 분산, 백업, 오류 검들 등의 관리 정책을 정형적으로 정의할 수 있다. 이로서 관리 정책을 자동적으로 수행할 수 있으며 관리 정책의 변경 및 추가에 유연하게 대처할 수 있다. 그리고 가상화 기능을 이용하여 분산되어 저장된 데이터들을 효율적으로 관리할 수 있으며 시스템의 확장성을 높일 수 있다.

본 논문에서는 규칙 기반의 그리드 시스템인 iRODS를 이용하여 전자 기록물 관리 시스템을 설계 및 구현하였다. 분산된 자원의 가상화를 통하여 시스템의 확장성을 높이고, 규칙을 이용하여 기록물 유형에 따라 데이터의 자동 분산, 백업, 복제된 데이터 오류 검사 등을 정의하여 저장 및 관리절차를 자동화할 수 있음을 보인다. 그리고 무결성 보장을 위하여 iRODS의 메타데이터를 관리하는 iCAT 서버를 이용하여 전자 기록물의 메타 데이터로 원본의 체크섬 값과 복제본의 개수 등을 저장한 후, 이를 이용하여 복제된 기록물과 비교하는 방법을 제시한다. 또한 국가 기록원의 기록물 관리 시스템을 모델링하여 구현함으로써 iRODS 기반 전자 기록물 관리 시스템이 실제로 어떻게 구현 및 적용되는지를 보인다.

본 논문의 구성은 다음과 같다. 2장에서는 전자 기록물 관리 시스템에 대한 관련 연구를 간략히 살펴보고, 3장에서는 iRODS의 기능들에 대하여 살펴본다. 그리고 4장에서는 iRODS를 이용한 전자 기록물 관리 시스템의

설계를 설명하고 5장에서는 이에 대한 실제 구현을 보여준다. 마지막으로 6장에서는 결론 및 향후 연구에 대하여 논의한다.

2. 관련연구

2.1 전자기록물 관리 표준 OAIS

전자기록물 관리 체계의 필요성이 대두됨에 따라 그에 따른 표준화를 추진하여 2002년 1월 국제 표준으로 전자기록물 관리 참조모형인 Open Archival Information System(OAIS)가 확정되었다. OAIS 참조모형은 장기간의 전자 기록 보존과 관리, 이용을 위한 기록 보존시스템으로서 디지털 아카이브를 구성을 위한 개념적인 구조를 정의하고 있다[7]. 이러한 OAIS 참조모형을 기반으로 전자 기록물 보존 아키텍처를 구축하고 보존 프로세스 워크플로우를 구성하여, 전자 기록물의 수집, 이관, 보존, 검색 등으로 구성된 전자기록 관리 시스템이 구축되고 있다.

그림 1에서 도식화한 OAIS 기본구조는 OAIS 참조모형의 큰 틀을 간단한 구조로 나타내고 있다. 이 기본구조는 OAIS 아카이브를 중심으로 Producer, Management, Consumer가 상호작용하는 환경에서 기능을 수행하는 구조이다. Producer는 보존 정보를 제공하는 사용자나 클라이언트 시스템 역할을 하고 Management는 OAIS의 전반적인 정책을 수립하는 역할을 한다. Consumer는 보존된 정보에 대한 검색과 열람을 요청하는 사용자나 클라이언트 시스템으로서 역할을 담당한다[3].

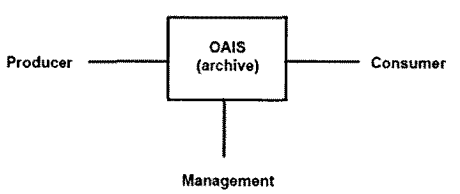

그림 1 OAIS 기본구조[8]

그림 2는 OAIS 기본 구조를 세부적인 6가지 기능별 영역으로 구분하고 Information Package의 역할에 대해 나타내고 있다[8].

- Ingest영역은 Producer로부터 입수기록패키지(Submission Information Package, SIP)를 접수받아 아카이브 저장하기 전 품질을 확인하며, 보존기록패키지(Archival Information Package, AIP)와 메타데이터인 기술 정보(Descriptive Information, DI)를 생성하는 기능을 한다.
- Archival storage영역은 AIP의 보존, 유지, 복구, 검색 등의 기능을 수행한다.

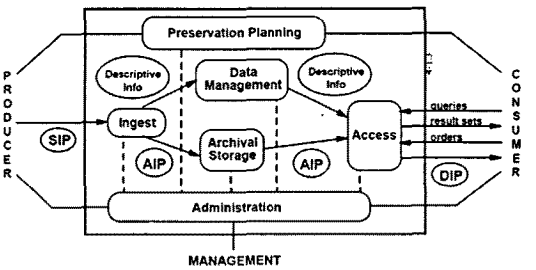

그림 2 OAIS의 6가지 기능영역과 Information Package[8]

- Data Management영역은 보존 정보에 대한 DI와 아카이브 운영에 필요한 행정정보에 대해 유지 및 접근 기능을 수행한다.
- Administration영역은 아카이브 시스템의 운영에 관한 기능을 수행한다.
- Preservation planning영역은 OAIS 환경을 감독하고 전산 환경 노후화에 대해 관리한다.
- Access영역은 Consumer가 정보의 존재 유무, 정보에 대한 설명, 정보의 위치, 정보 입수 가능 여부 확인, 정보 산출물(Dissemination Information Package, DIP)에 대한 요청을 하면 이를 접수 및 처리하고 요청된 DIP를 입수하여 Consumer에게 전달하는 역할을 한다.

전자기록물 보존·관리 시스템은 OAIS 표준 모형을 기반으로 전자기록물 보존·관리를 위한 시스템 표준화를 통해서 전자기록물의 장기보존에 적합한 시스템으로 구축되어야 한다.

2.2 전자기록물 관리를 위한 자동화 연구

지속적으로 증가하는 대량·대용량의 전자기록물의 관리에 있어서 수동적인 업무 프로세스에 의해 발생하는 노동력 낭비 및 고비용성의 문제, 비효율성 등을 해결하기 위하여 보존 및 관리 과정의 자동화가 이루어져야 한다. 또한 전자 기록물 저장소 확장에 따른 확장성 및 호환성을 확보하고 전자기록물 발생 기관별 정책적 특성과 상황적인 특성을 반영할 수 있는 전자기록 자동화 관리 시스템이 요구된다. 유용한 전자기록 자동화 관리 시스템의 구축을 위해서는 이러한 확장성, 호환성, 유연성 등의 요구사항을 만족시키는 자동화 프로세스를 도입하여 야 한다. 그림 3은 기존 자동화 프로세스의 문제점을 도식화하였다. 예를 들어 기존에 자동화 시스템에 새로운 프로세스 체계로 추가·변경 시에 기존의 구축된 시스템의 유연성 부족으로 인해 변경에 어려움이 발생한다. 또한 기존 자동화 시스템에 새로운 자동화 시스템을 추가하고자 할 때 이기종 시스템과의 호환성 문제가 발생한다.

2.3 전자기록물 관리를 위한 분산 처리 연구

분산 처리는 네트워크를 이용하여 분산된 시스템의

자원 공유와 병렬 처리를 수행하는 것을 의미한다. 분산 처리는 부하분산, 처리분산, 기능분산, 위험분산, 관리분산, 확장분산 등의 다양한 기능적 특성을 갖는다. 이러한 분산 처리는 대량·대용량의 전자기록물을 저장 및 관리하고, 전자기록물의 무결성, 신뢰성을 유지하기에 적합한 특성을 갖추고 있다. 또한 전자기록물을 지리적으로 분산 관리하여 재해와 재난을 대비하고 보존 시스템 확장의 위치적 공간적 한계를 극복하는 해결책이 될 수 있다.

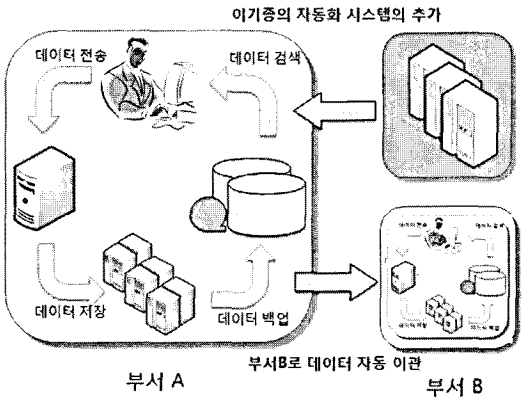

그림 3 기존 자동화 시스템의 호환성/유연성 문제

그리드 컴퓨팅[9,10]은 지역적으로 분산된 다수의 이기종 컴퓨팅 자원을 네트워크로 연동하여 가상의 대용량 고성능 컴퓨팅 시스템을 구성하는 분산 컴퓨팅 기술이다. 그리드 기술 중 데이터 분산 처리 및 저장, 공유를 목적으로 구성된 그리드 컴퓨팅을 데이터 그리드라고 한다. 데이터그리드는 분산된 컴퓨팅 자원을 네트워크로 연동하여 거대한 저장소를 구성하여 데이터 공유와 저장을 목적으로 사용된다. 현재 다양한 분야에서 데이터 공유, 디지털 라이브러리, 디지털 아카이브 구축 등에 활용되고 있다. 이러한 데이터 그리드는 다양한 분산 컴퓨팅 기술 중에서 대량·대용량 데이터 처리를 위해서 가장 적합한 특성을 지니고 있다. 표 1은 현재 데이터 그리드를 이용한 프로젝트들이다.

2.4 대량·대용량 데이터 규모 적응성

데이터의 규모 적응성이란 대량·대용량으로 데이터가 발생되는 환경에서 시스템적 유연성과 확장성을 보장하는 특성을 의미한다. 전자기록물 관리를 위한 시스템 구축은 시간이 지남에 따라 저장 공간이 한계에 다다르며, 추가적인 시스템의 확장이 요구되어진다. 추가적인 시스템의 확장 시, 이미 구축된 시스템과 이미 저장된 자료에 영향이 없이 계속적인 운영을 위해서는 초기 설계 시 데이터 규모 적응성을 고려하여야 한다. 데이터 그리드 기술은 이러한 요구를 가상화 기술을 통하여 충족시킨다. 데이터 그리드 기술은 물리적인 저장 공간을 가상화하여 그리드에 참여하는 시스템의 종류에 독립적이며, 데이터 저장과 관리에 있어 높은 확장성과 호환성을 보장한다. 이를 통해 데이터 저장 공간 부족 문제와 시스템 확장 시의 호환성 문제를 해결함으로써 대량 데이터에 대한 규모 적응성을 갖는 시스템 구축이 가능하다.

표 1 데이터 그리드 프로젝트

프로젝트 목적	데이터 그리드 프로젝트
생물학 연구	Biomedical Informatics Research Network (BIRN)
지질학 연구	Southern California Earthquake Center (SCEC)
환경 변화 연구	Real-time Observatories, Applications and Data management Network (ROADNet)
보존 관리	Digital Libraries and Archives: National Science Digital Library (NSDL)
	National Archives and Library of Congress
	Univ. of Michigan Digital Library Archive

2.5 iRODS를 이용한 저장 시스템

iRODS는 2008년 1.0 버전이 발표된 이후 저장 시스템의 구축 및 다양한 프로젝트에 적용되어왔다. 저장 시스템에 대하여 적용된 대표적 프로젝트는 다음과 같다[11].

- Time for Learning : New Scientific Insights to Educators
- Virtual Earthquakes Help Prepare for "the Big One"
- "Climate Crisis" in Western U.S. More Certain
- Bringing Michigan Election Data Online
- iRODS Opens Virtual Vistas for Astronomy
- Speeding Information to Help Predict "the Big One"
- Dialing for Data: iRODS Collaborations Connect Science

이와 같이 iRODS는 대량·대용량 데이터들의 저장 시스템 구축에 적용되어 성능은 검증되었으나, 각 프로젝트들에서 사용된 설계방법이나 정책을 표현한 규칙들에 대한 문서화가 되어 있지 않다. 그러므로 iRODS를 이용하여 저장 시스템을 구축하기 위한 모델링 방법과 규칙의 적용 방법 등이 필요하다.

3. 규칙기반 데이터 그리드 시스템: iRODS

iRODS는 규칙 기반 데이터 그리드 시스템으로서 지리적으로 분산된 컴퓨팅 자원을 네트워크로 연동하여 가상화를 구성하고 이를 통해 대용량 저장인프라를 구성하는 기술이다. iRODS의 대표적인 기능은 가상화를

통한 데이터 분산 처리 및 이기종 시스템과의 호환성 보장, iCAT(iRODS Catalog)을 이용한 메타데이터 관리, 규칙(rule)을 이용한 가상화 정책 관리 및 자동화이다. 가상화 기능은 지리적, 물리적으로 분산된 시스템 자원을 하나의 논리적인 공간으로 설정하여 전체 시스템 관리의 편의성을 제공하고 시스템 확장에 따른 유지보수성의 문제를 해결하며 이기종 시스템과의 호환성을 확보한다. iCAT은 보존 기록에 대한 메타데이터를 관리하여 기록의 검색과 복구에 대한 처리가 가능하다. iRODS는 기록물에 대한 분산 및 백업 정책 등과 같은 OAIS 표준을 따르기 위한 여러 프로세스들을 정의하고 관리하기 편리한 정책(규칙) 기반의 프로그램을 지원한다. 자료 관리를 위한 정책들은 복제 및 복사, 자료의 유효성을 검증하는 등의 작은 단위 작업을 수행하는 마이크로 서비스들을 조합하여 규칙이라는 일련의 작업들의 집합으로 프로그램 된다. 이러한 규칙은 가상화된 공간에서 데이터 관리 및 정책을 수립하여 자료 관리에 대하여 사용자가 정의할 수 있는 조작 환경을 제공한다. 이와 같이 iRODS는 전자기록물 관리에서 요구하는 확장성과 호환성, 일관성, 자동화 기능들을 제공함으로써 전자기록물 관리에 효과적으로 사용될 수 있는 규칙 기반의 데이터 그리드 시스템이다.

그림 4는 iRODS 데이터 시스템의 구조를 도식화 한 것이다. 사용자들은 사용자 인터페이스(Web or GUI Client)를 통해서 iRODS 시스템에 접근하여 파일을 저장 및 검색할 수 있다. iRODS 시스템은 파일 저장 요청이 발생하면 파일의 메타데이터를 iCAT(iRODS Metadata Catalog) DB에 저장하고, 파일을 규칙 엔진의 규칙에 의해 분산된 iRODS 서버에 저장한다. 그리고 검색 요청이 발생하면 파일에 대한 메타데이터를 iCAT DB에서 찾아 해당하는 파일을 분산된 iRODS 서버로부터 검색한다.

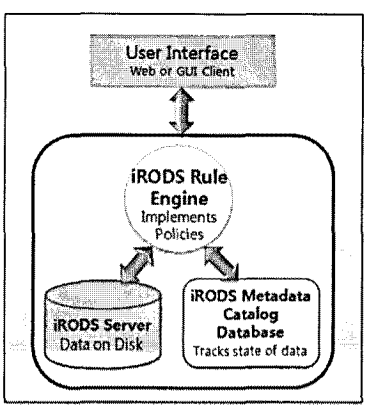

그림 4 iRODS 시스템 구성도

규칙 엔진은 관리자에 의해 사전에 규칙을 정의하거나 실행시간에 규칙을 적용하여 분산된 데이터 저장서버(iRODS Server)에 파일을 저장하고 iCAT DB에 저장물의 기본적인 메타 데이터 및 사용자 정의 메타 데이터를 저장한다.

3.1 규칙 기반 자동화

전자기록물의 관리에 있어 수동적인 프로세스에 의해 발생할 수 있는 비용과 오류, 비효율을 개선하기 위하여 자동화가 요구된다. 전자기록물 관리에 있어서 필요한 절차를 iRODS가 제공하는 파일의 분산, 복제 등의 단위기능들(마이크로서비스)을 순차적으로 조합하여 만들어진 규칙(일련의 수행 정책)을 구현함으로써 자동화를 구성할 수 있다. 이러한 구성방법은 규칙 설정 파일을 스크립트 수준의 추가 및 변경을 통하여 수행할 수 있기 때문에 시스템의 유연성을 보장할 수 있다.

3.1.1 규칙 정의

규칙은 매크로 단계 작업(macro-level task)들 또는 액션(action)이라 불리는 단위기능들(마이크로 서비스)들의 조합을 통해 정의한다. 마이크로 서비스는 데이터를 다루거나 데이터의 유효성을 검사하는 기능적 역할을 제공한다. 표 2는 마이크로 서비스의 대표적인 예를 나타낸다.

표 2 마이크로 서비스의 예

마이크로서비스	동 작
msiDataObjOpen	데이터 객체 열기
msiDataObjClose	열려있던 데이터 객체 닫기
msiDataObjRead	열려있는 데이터 객체 읽기
msidataObjWrite	데이터 객체 상에 쓰기

규칙은 다음과 같이 규칙명칭, 실행조건, 마이크로서비스, 복구명령으로 구성된다. 규칙 내의 각 요소들은 "|"으로 구분하며, 마이크로 서비스들은 "##"를 통해서 구분하여 연결한다.

Rule ↦ 규칙명칭(actionDef) | 실행조건(condition) | 워크플로우체인(workflow-chain) | 복구명령(recovery-chain)

- actionDef: 규칙의 이름을 정의한다.
- condition: 규칙이 실행되기 위한 조건을 정의한다. 예를 들어 "condition ↦ $objPath like /x/y/z/*" 는 x/y/z/경로의 콜렉션(논리적 경로 상의 디렉토리를 의미)에 데이터 객체가 생성되면 규칙을 실행하는 실행 조건이다.
- workflow-chain: 조건에 부합되는 상황이 발생되면 실제로 동작하게 되는 마이크로 서비스들 또는 규칙의 연결로써 '##'의 형태로 연결되어 순차적으로 실행되는 구조를 갖는다. 하나의 규칙은 여러 개의 마이

크로서비스나 규칙들로 구성될 수 있다.

- recovery-chain : 워크프로우 체인의 실행 후 실패한 워크프로우 체인에 대한 복구를 담당한다. 복구는 앞서 선언된 액션의 순서와 같은 구조를 띤다.

3.1.2 규칙 등록

정의한 규칙들은 시스템레벨과 유저레벨로 시스템에 등록되어 그 역할을 수행할 수 있다. 시스템레벨 규칙은 인증, 무결성, 접속의 제한, 데이터의 복제, 분산 등 데이터관리정책등을 포함한다. 유저레벨 규칙은 클라이언트가 API를 통하여 외부에서 지정하여 실행하는 방식으로 주로 작업들을 순차적으로 실행하도록 하는 워크플로우 형태로 나타난다. 유저레벨 규칙은 분산된 서버들 각각의 시스템레벨 규칙에 독립적으로 적용되므로, 서비스 제공시 일관성 있는 규칙의 적용에 매우 효과적이다[4].

규칙의 등록은 파일의 확장자를 .irb인 텍스트 파일로 작성하여 iRODS 설치경로상의 디렉토리인 *server/config/reConfigs* 에 저장한 후, 설치경로상의 *server/config/server.config* 의 파일의 reRuleSet이라는 환경변수를 수정함으로써 등록된다. 환경변수의 값은 분리자(,)로 분리되어지며, 하나이상의 규칙 파일을 등록할 수 있고, 기본적으로 core.irb 파일이 지정되어 있다[4].

유저레벨 규칙은 실행시간에 사용자가 irule 명령이나 rcExecMyRule API를 호출함으로써 이루어진다.

각 규칙의 등록 방법은 분산된 서버들 각각의 로컬 정책 등은 시스템 레벨 규칙을 적용하여 개별적인 데이터의 백업이나 인증 등을 수행하도록 하고, 전체적인 서비스에 관한 워크플로우등은 제공하는 서비스에서 유저레벨 규칙을 적용함으로써 유연성을 확보할 수 있다.

3.2 가상화를 이용한 분산 처리

분산 처리는 대량·대용량으로 발생하는 전자기록물 처리에 있어서 확장성과 호환성을 보장한다. iRODS는 지리적으로 구애받지 않고 가상화된 분산 처리 시스템을 구축하여 확장성과 호환성 등의 기록물 관리 시스템으로서 필요조건을 만족하는 환경 구성이 가능하다.

iRODS를 이용한 분산처리는 저장 공간 가상화, 리소스, 리소스 그룹화 등을 통해서 이루어진다. 가상화는 다수의 물리적인 공간을 논리적인 단일 공간으로 바꾸어, 논리적인 단일 공간을 이용할 때 자동적으로 다수의 물리적 공간에 접근하여 데이터를 저장, 관리할 수 있도록 해준다. iRODS는 물리적 저장 공간인 컴퓨팅 자원을 리소스(resource)형태로 지정하여 가상화된 논리 공간(zone)과 연결시킨다. 따라서 사용자는 가상화된 공간에서 다양한 리소스를 선택하여 물리적 데이터 분산 처리를 수행하게 된다. 또한 iRODS는 단일 리소스를 사용한 저장뿐만 아니라 리소스를 가상 그룹화 하여 여러 리소스들을 하나로 묶어 관리하는 것이 가능하다. 그룹화 기능을 통해 여러 리소스들을 하나의 그룹으로 관리함으로써 새로운 시스템의 도입 시에 하나의 시스템 리소스 그룹으로서 인식하여 기존 시스템의 변경 없이 시스템의 확장성을 보장한다. 예를 들면, 배포하는 서비스/응용프로그램에서 단일 리소스를 사용하여 저장 때 해당 리소스가 한계에 도달하였을 경우, 시스템의 재구성이나 서비스/응용프로그램의 재배포가 이루어져야 하지만, 가상 그룹화 된 리소스 그룹을 사용하여 저장하도록 할 경우, 서비스/응용프로그램의 재배포의 필요 없이 리소스 그룹에 또 다른 리소스를 추가하는 것으로 문제를 해결할 수 있다. 이러한 리소스 그룹을 사용할 경우, 그룹 내의 리소스 분배 등의 정책은 규칙의 등록을 통하여 유연하게 변경 가능하다.

3.3 분산 데이터의 무결성

기록되는 데이터가 보존을 위하여 복제된 후, 분산되어 보관될 경우에 원본과 복제본간의 무결성이 보장되어야한다. 기록물 보존에 있어서 무결성은 데이터 무결성과 시스템 무결성으로 나눠진다. 데이터 무결성은 규정되고 인가된 방법으로만 데이터가 변경되어야 함을 의미하고, 시스템 무결성은 고의로 또는 부주의에 의한 허락되지 않은 조작으로부터 시스템이 손상되지 않고 정상적인 기능을 발휘함을 의미한다[12]. 이러한 무결성은 적절한 규칙을 적용함으로써 보장이 될 수 있다. 적절한 규칙을 적용한다는 의미는 일련의 복제, 분산의 과정을 자동화하는 규칙을 정의하여 최초 입력에 따라 그 외의 과정을 시스템에서 관리함으로써 규정되지 않고 비인가된 방법으로의 보호를 의미한다.

이러한 규칙의 적용을 위해 iRODS는 Delayed Execution Service와 단위기능 중 체크섬을 확인하는 마이크로서비스를 지원한다. iRODS는 규칙을 적용할 때, 기록물의 저장 요청 전과 후로 나누어 적용을 하게 되는데, 원본과 복제본 사이의 무결성을 검사하기위해서는 기록물이 저장되고 복제된 후 분산이 완료된 시점에서 체크섬을 확인하여야 한다. Delayed Execution Service는 기록물의 분산이 완료된 시점에 규칙을 적용할 수 있도록, 규칙과 단위기능들을 큐잉 하였다가 실행하도록 한다[4].

복제 데이터의 분산 시에 무결성은 기록물 저장 과정을 자동화하는 규칙과 iRODS의 체크섬을 확인함으로써 보장할 수 있으며, 이러한 무결성에 대한 정보는 iRODS의 구성 중에 하나인 iCAT DB에 저장된다. 저장된 기록물을 검색하여 재 호출할 경우, iRODS는 iCAT DB에 저장된 체크섬 값으로 복제본의 진위여부를 판별한다.

3.4 대량·대용량 데이터 전송의 효율성 향상을 위한 방법

데이터를 iRODS 시스템에 저장하기 위해 iPUT이라는 명령을 사용한다. iPUT 명령은 대량의 데이터를 처리하기 위해 멀티 쓰레드를 지원하도록 설계되었으며, 대용량의 데이터를 처리하기 위해 하나의 iPUT 명령은 파일 크기에 따라 여러 개의 쓰레드로 나누어 동작하며, 최대 16개의 쓰레드를 사용하도록 되어있다. 예를 들어 32MB 이하의 데이터의 전송 시에는 한 개의 쓰레드를 사용하고, 32MB에서 63MB사이의 데이터는 두 개의 쓰레드를 사용하며, 512MB 이상의 자료는 16개의 쓰레드를 사용하게 된다. 생성하는 쓰레드의 개수를 통하여 성능을 향상할 수도 있지만, iRODS 클라이언트/서버간의 슬라이딩 윈도우 사이즈를 조절하여 성능을 향상 시킬 수 있다. 각각을 위해 iRODS는 설정 값을 조절할 수 있는 방법을 제공하고 있으며, 대용량 파일 전송 시 iPut 당 2개의 쓰레드를 할당하고, 16MB의 슬라이딩 윈도우를 설정할 때 가장 좋은 성능을 낸다[13].

4. iRODS를 이용한 전자기록물 관리 시스템 설계

그림 5는 전자 기록물 관리 시스템의 예로서 국가 전자 기록물의 보존 절차를 도식화 한 것이다. 각 부서는 기관별 관리 정책에 따라 전자 기록물을 국가 기록원의 수집 서버에 저장하며 데이터의 유형에 따라 분류되어 대전, 부산 등에 산재된 데이터 저장소에 자동적으로 분산/복제 된다. 그리고 각 데이터에 대한 정보는 메타데이터로서 관리가 되며 이를 통하여 기록물의 검색 기능을 제공한다. iRODS는 규칙을 이용한 자동화, 분산처리, 메타데이터 관리 등의 기능을 이용하여 이와 같은 전자 기록물에 대한 수집/저장/검색 등의 관리 시스템을 효과적으로 구축할 수 있다.

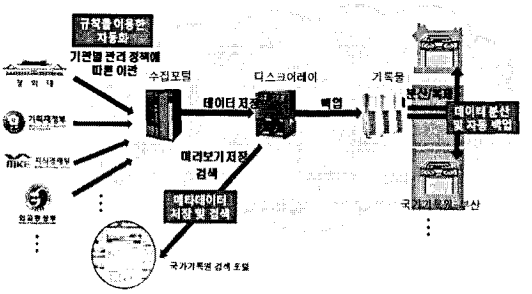

그림 5 국가 전자기록 보존 절차

4.1 분산된 자원의 가상화 설계

iRODS는 분산된 물리적 저장 공간을 논리적 저장 구조로 다루기 때문에 대용량 전자 기록물 관리를 위해서는 논리적 저장 구조에 대한 유용한 설계방법이 우선시 되어야 한다. 이를 위해서 iRODS에서는 Zone과 Resource와 Collection이라는 요소를 지원하며, Resouce와 Collection의 구조를 설계하여 구성하고 기록물에 대하여 규칙을 적용함으로써 대용량 전자 기록물을 관리한다. Zone은 하나의 iRODS 영역과 대응하며, 하나의 Zone은 계층적 구조의 Collection을 갖게 된다. 현재 iRODS를 이용하여 하나의 Zone만을 관리할 수 있으며, 추후 여러 Zone을 관리할 수 있도록 확장될 것이다. Resource는 분산된 물리적 저장 공간을 논리적으로 관리하는 요소이다. 현재 iRODS에서 지원하는 Resource의 형태는 unix file system, HPSS, Amazone S3에 한정되며, 추후 계속 늘어날 것이다. 여러 개의 Resource는 Resource Group으로 그룹화 될 수 있으며, 이러한 그룹화는 저장되는 기록물의 증가에 따른 저장공간의 규모 가변성을 보장할 수 있다. 예를 들어, Resource Group의 저장 공간이 모두 사용되었을 때, Resource Group에 추가적인 Resouce를 할당하는 것만으로 연결되어있는 다른 규칙이나, 시스템에 영향을 미치지 않고 본래의 기능을 계속적으로 수행할 수 있다.

Collection은 iRODS를 구성하는 가상화된 파일시스템의 디렉토리 역할을 하는 요소이다. iRODS에 기록물이 저장될 경우, Resource와 Collection을 지정하게 되는데, Resource를 지정함으로써 물리적인 분산위치를 결정하고, Collection을 지정함으로써 논리적인 분산위치를 결정하게 된다.

전자기록물 관리 시스템의 설계에 있어서, Resource를 할당하는 방법으로는 시스템을 사용하는 조직의 구성에 대응하도록 구성하는 방법과 시스템을 구성하는 지리적 위치에 대응되도록 구성하는 방법으로 나눌 수 있다. 그림 6은 각 구성방법을 도식화한 것이다. 조직의 구성에 대응하도록 Resource를 구성하였을 경우의 이점은 물리적인 저장장치들이 지리적으로 분산되지 않고 물리적 저장장치를 공유하여 사용할 때, 물리적 저장장치를 논리적으로 분할하여 관리할 수 있다. 그러므로 기록물을 조직 구성원간의 분산이나 복제 등을 통하여 공유할 경우 유용하다. 지리적 위치에 대응하여 Resource를 구성할 경우의 이점은 대규모의 정부 기관이나 기업

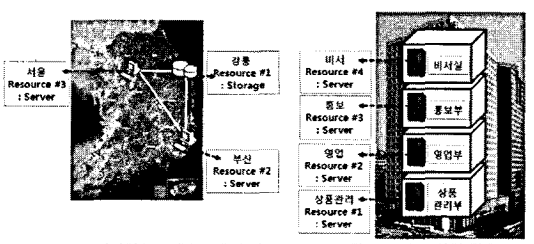

그림 6 Resource의 구성 예

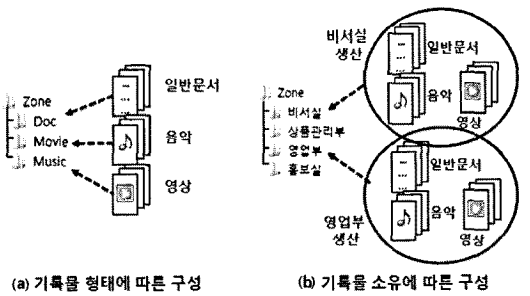

그림 7 Collection의 구성 예

등에서 분산된 물리적 자원들을 중앙에서 효율적으로 관리하며, 지리적으로 분산된 저장 자원들에 대하여 백업이나 분산을 통하여 재난 복구 등의 목적에 부합시키고자 할 경우 유용하다.

Collection의 구성에 있어서 기록물의 형태(즉, 자료의 성격)에 따라서 구성을 하는 방법과 기록물의 소유자에 따라 Collection을 구성하는 방법이 있다. 그림 7은 각 구성방법을 도식화한 것이다. 전자의 경우, 기록물의 종류가 중요시 되는 경우, 물리적 저장위치와 관계없이 기록물의 성격에 따라 자료를 분산하고자 할 경우 유용하며, 후자의 경우 시스템을 사용하는 조직별로 자료가 관리되어야 할 경우 유용하다.

4.2 iRODS를 이용한 업무처리의 자동화 설계

4.2.1 iRODS를 이용한 기록물 유형에 따른 데이터 자동분산설계

iRODS에서의 분산은 기록물 유형에 따라 여러 Collection으로 분산하여 저장하는 방법과 여러 Resource로 분산하여 저장하는 방법으로 나눌 수 있다. 여러 Collection으로 분산한다는 의미는 논리적으로 구분되어 있지만 물리적 위치는 고려하지 않고 분산저장을 하는 것이고, 여러 Resource로 분산한다는 의미는 특정 물리적 위치를 고려하여 분산저장을 하는 것이다.

이러한 자동 분산의 실제적인 매개체(실행자)는 3절에서 설명한 iRODS의 마이크로 서비스를 이용하게 되며, 실행 시간에 응용단에서 기록물의 메타데이터를 확인하여 Resource나 Collection을 지정하여 마이크로 서비스를 호출하거나, 규칙을 통하여 메타데이터를 확인하고 Resource나 Collection을 지정하게 된다. 전자의 경우, iRODS에서 제공하는 클라이언트 API를 이용하여 응용프로그램을 설계/제작하는 경우 유용하며, 후자의 경우 iRODS에서 제공하는 범용적인 클라이언트 프로그램을 사용할 경우 유용하다.

4.2.2 iRODS를 이용한 데이터 자동 백업

iRODS를 이용한 백업의 경우, 해당 데이터를 원본에 관계없이 저장하는 방법과 재난에 대비하여 저장하는 것으로 나눌 수 있다. 전자의 경우, 데이터 원본을 사용자가 삭제하더라도, 데이터가 유지되며, 이러한 데이터 원본을 별도의 저장소에 보관되도록 데이터를 복사하는 마이크로 서비스를 이용하게 된다. 후자의 경우, 데이터 원본을 사용자가 삭제 시에 백업된 데이터도 같이 삭제되어진다. 후자의 경우는 데이터를 복사하는 마이크로 서비스가 아닌 복제하는 마이크로 서비스를 사용하게 되며, 전자와 후자의 차이는 메타 데이터 공유의 차이를 가진다. 전자의 경우는 메타데이터가 공유되지 않으며, 이 경우, 데이터의 무결성의 문제가 생길 수 있다. 이러한 무결성의 문제를 해결하기 위해서 추가적인 연산이 필요하게 된다. 후자의 경우는 메타 데이터를 공유하며, 데이터의 무결성을 보장한다. 백업에 대한 각각의 선택은 설계하고자하는 전자기록물 관리시스템의 백업 요건에 따른다.

4.2.3 iRODS를 이용한 메타 데이터 저장 및 검색

iRODS는 iCAT DB를 이용하여 메타 데이터를 관리하며, 메타 데이터는 기록물에 대하여 iCAT DB내부에서 기본적으로 관리하는 메타 데이터와 사용자가 필요에 의해 정의하는 사용자 정의 메타 데이터로 나뉜다. iRODS에서 입력된 기록물에 대하여 제공하는 기본적인 메타 데이터는 다음과 같다.

1.이름
2.크기
3.소유자
4.소유자가 속한 Zone
5.최종 수정 일자
6.생성일자
7.아이디(기록물의 참조 인덱스)
8.저장된 리소스 이름
9.설명
10.복제된 기록물의 개수

기본적인 메타 데이터 외에 기록물에 대한 추가적인 정보를 설정하기 위해서는 사용자 정의 메타 데이터를 이용하여야 한다. 사용자 정의 메타 데이터는 기록물을 iRODS 상에 입력 시 런타임 상에서 스트림 또는 파일로 입력된 데이터를 통하여 기록물에 대한 참조 값을 가지고 iCAT DB상에 입력된다. 이렇게 입력된 데이터는 클라이언트 API를 통하여 기록물의 검색 및 정보 확인에 이용되어 진다.

각 관리방법에 대한 요구를 만족하기 위해서는 사용자 정의 메타 데이터를 정의하여야 한다. 이러한 사용자 정의 메타 데이터는 응용프로그램을 설계하는 당시에 기록물에 대해 부가적으로 입력되는 정보들이 주로 사용된다. 사용자 정의 메타 데이터는 응용프로그램 실행 시 런타임 상에서 iCAT DB로 입력되며, 사용자 정의

메타 데이터는 iRODS에 입력되는 전자기록물과 연결되어있다. 사용자 정의 메타 데이터는 iRODS를 이용하는 응용프로그램에서 iRODS에 파일을 요구할 경우 검색 조건으로 사용되며, 해당 전자기록물의 세부 사항을 설명하는데 사용된다.

5장에서는 위의 설계 방법 및 메타 데이터를 이용한 실제 테스트베드를 구현함으로써 iRODS를 이용한 전자기록물 관리 시스템 구축의 예를 보인다.

5. iRODS를 이용한 기록 관리 시스템의 구현

본 절에서는 3절에서 소개한 iRODS와 4절에서 소개한 기본적인 설계 방향을 이용하여 국가기록물에 대한 전자기록물 관리 시스템을 모델링하여 구현하였다. 현재 국가기록원에서의 기록물 관리는 정부산하 700여 부서에서 10년 이상 된 문건들에 대하여 국가기록원으로 이관하며, 이관되는 문서는 성남에서 저장하며, 네트워크 사용율이 낮은 야간시간대에 대전 본원에 복제 및 저장을 함으로써 이원화하여 관리하고 있다. 이러한 현재 시스템에서는 원본과 복제본 간의 무결성의 확인이 어려우며 동시성을 보장할 수 없다. 그리고 분산된 자원들은 개별적으로 관리가 되기 때문에 일괄적인 관리정책을 적용할 수 없다. 이러한 문제점을 해결하기 위해 본 논문에서 저장시스템의 프로토타입을 구현하였으며, 저장시스템이 갖추어야 할 고려사항으로 자원의 가상화를 통한 시스템의 규모적응성과 유지보수성을 향상시키고, 자료 처리의 과정을 규칙을 이용하여 자동화 함으로써 사용자의 실수나 오작동으로 인한 자료의 안전성을 향상시켰다. 구현된 시스템은 대량·대용량 전자 기록물에 대하여 규칙을 이용한 분산 및 복제를 수행함으로써 자동화를 지원하고, 지역적으로 분산된 자원을 가상화를 통하여 논리적으로 통합하여, 시스템의 확장성을 높이며, 분산되어 복제된 데이터에 대한 정당성을 확인함으로써 무결성을 보장한다.

5.1 국가기록물에 대한 전자 기록물 관리 시스템의 모델링

국가기록물에 대한 전자 기록물 관리 시스템은 다음의 요구사항을 만족하여야 한다고 가정한다.

1. 지리적으로 대전, 성남, 부산에 분산된 3곳의 기록소로 구성되어 있다.
2. 국가기록물은 각 정부 부서별로 음성 및 동영상 기록물, 필름 및 사진 기록물, 일반 문서 기록물 등이 생산 되며, 기록물이 생산 부서별로 저장되는 것보다 기록물의 유형별로 지역적으로 분산시키기를 원한다.
3. 모든 기록물은 기록물 철, 기록물 건, 기록물로 정의되는 계층형 구조를 가진다. 기록물철은 여러개의 기록물 건으로 구성되며, 기록물 건은 여러개의 기록물을 파일로서 가지게 된다.
4. 기록물은 해당 기록물 철의 기록유형에 따라 다음과 같은 물리적 위치 지정 정책에 따라 저장한다.
   - 일반문서, 도면류, 카드류 일반기록물: 대전 본원
   - 녹음 및 동영상류 시청각기록물: 성남 나라 기록관
   - 사진 및 필름류 시청각기록물: 부산 역사 기록관
5. 백업 정책은 다음과 같다.
   - 물리적으로 대전 본원과 성남 나라 기록관, 부산 역사 기록관에 저장되는 임의의 데이터에 대하여 성남 나라 기록관 백업장치에 저장된다.
   - 저장되는 논리적 위치는 해당 자료가 속하는 기록물철의 기록유형에 따라 해당 위치에 복제(replication)된다.

이러한 요구 사항을 충족시키기 위해 그림 8과 같이 iRODS의 Resource와 Collection을 구성하였다. 대전 및 성남, 부산에 분산되어 있는 물리적 저장소를 PC1, PC2, 외장하드1, 외장하드2로 대응시키고, 리소스 명을 pc1, pc2, ext_hdd1, ext_hdd2로 지정하였다. 또한, 논리적 분산을 위해 Doc, Audio-Video, Film-photo로 Collection을 생성하였다.

그림 9는 iRODS를 이용한 전자문서 관리 시스템에서의 서비스를 표현하는 use case 다이어그램이다. 행위자는 기록자와 검색자로 나뉘며, 기록자는 파일 및 메타데이터를 전송하고 시스템은 파일을 기록유형에 따라 파일을 분산 저장하고 복제정책에 따라 복제하고 메타데이터를 iCAT DB에 저장한다. 검색자는 검색어를 시스템에 전송하고, 시스템은 이 검색어를 이용하여 iCAT DB가 관리하는 메타데이터를 검색하여 그에 맞는 기록물과 정보를 검색자에게 반환하게 된다.

요구사항의 가정에 따른 시스템의 구성과 기록 프로세스는 그림 10과 같다. 전체 시스템 구성은 2대의 PC와 2대의 외장하드디스크로 구성하였고, 우분투 리눅스를 운영체제로 사용하였다. 1대의 PC에 iRODS Server와 iCAT DB를 설치하여 메인 서버로 구성하였고 다른 PC에 iRODS Server를 설치하여 메인서버와 연결하였

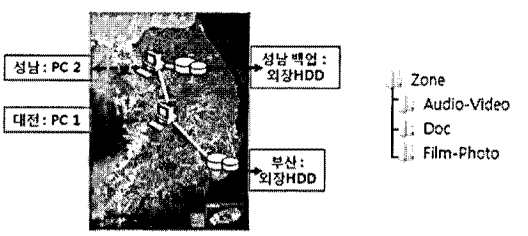

그림 8 국가기록원에 대한 Resource 및 Collection 모델링

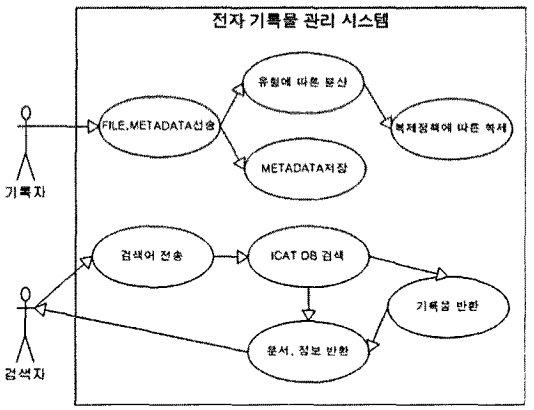

그림 9 전자문서 관리 시스템의 use case 다이어그램

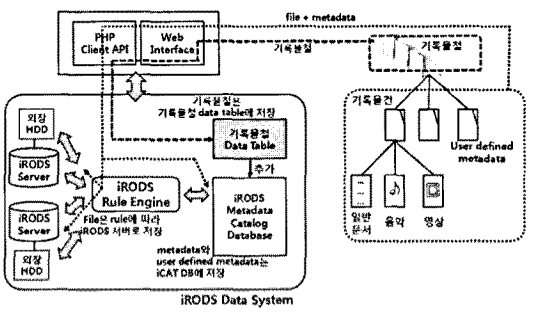

그림 10 기록물 관리시스템 구성

다. 각 PC에 외장하드디스크를 연결하였고, 2대의 PC와 2개의 외장하드디스크는 그림 9에 나타난 것과 같이 iRODS상의 리소스로 지정하였다. 이러한 분산된 시스템을 논리적으로 통합함으로써 분산된 자원의 가상화를 구현하였으며, 시스템의 확장 시 추가되는 리소스를 가상화된 논리공간에 추가함으로써 시스템의 유연한 확장성을 보장한다.

iRODS 시스템을 이용하기 위하여 iRODS에서 제공하는 웹스크립트 인터페이스인 php API를 사용하여 유저인터페이스를 구성하였다.

요구사항의 가정에 따른 자료의 계층적 구조는 그림 10에 표현된 것과 같이 기록물 철-기록물건-기록물의 계층 구조를 가진다. 기록물 철은 iCAT DB에 별도의 테이블을 생성하여 관리한다. 기록물건과 기록물에 대한 정보는 메타 데이터로서 iCAT DB에 저장되어 관리되며, 실제 기록물인 시청각기록물은 유저 인터페이스를 통해 웹서버에 저장된 후 해당 기록물의 메타 데이터를 분석하여 기록물의 유형에 따라 런타임 상에서 정책을 선택하고, 선택된 정책에 따른 규칙이 iRODS Rule Engine에 전달된다. 전달된 규칙에 따라 해당 기록물은 적절한 위치에 분산 및 복제 과정을 거치게 된다. 구현된 시스템은 정책을 선택할 수 있도록 구성되었으며, 다음과 같은 정책들을 정의하여 클라이언트 레벨로 규칙을 실행하도록 하였다. 의미적으로 표현하였으며, 해당 작업은 그에 대응하는 마이크로 서비스에 해당한다[14].

- 정책 1 - 저장

```
ingestInCollection(S):-
        chkCondition(S),ingest(S),register(S)
```

저장만을 수행하며, S는 입력파일을 의미한다.
입력된 파일은 chkCondition 단계에서 파일의 유형을 검사한 후, ingest 단계에서 원격지의 파일을 찾아서, register 단계에서 iCAT DB에 메타데이터를 등록하고, 파일 S의 유형에 맞는 위치에 파일 S를 저장한다.

- 정책 2 - 저장, 백업

```
ingestInCollection(S):-
        chkCondition(S),ingest(S),register(S)
        ,findBackUpRsrc(S,R),replicate(S,R)
```

정책 1의 과정과 같은 과정으로 register 단계까지 실행한 후에 findBackUpRsrc 단계에서 백업할 위치를 찾아서 R로 반환하고, replicate 단계에서 R에 S를 복제한다.

- 정책 3 - 저장, 검사

```
ingestInCollection(S):-
        chkCondition(S)
        ,computeClientChkSum(S,C1)
        ,ingest(S),register(S)
        ,computeServerChkSum(S,C2)
        ,checkAndRegisterChkSum(C1,C2,S)
```

chkCondition 단계에서 파일의 유형을 검사한 후, computeClientChkSum단계에서 입력파일의 체크섬을 C1으로 반환하고, ingest와 register 단계를 수행한다. computeServerChkSum 단계에서 저장된 파일의 체크섬을 C2로 반환하고 checkAndRegisterChkSum 단계에서 C1과 C2의 체크섬을 비교 후 결과를 저장한다.

- 정책 4 - 저장, 검사, 백업 & 검사

```
ingestInCollection(S):-
        chkCondition(S)
        ,computeClientChkSum(S,C1)
        ,ingest(S),register(S)
        ,computeServerChkSum(S,C2)
        ,checkAndRegisterChkSum(C1,C2,S)
        ,findBackUpRsrc(S,R),replicate(S,R)
        ,computeServerChkSum(S,C3)
        ,checkAndRegisterChkSum(C2,C3,S)
```

정책 3의 과정과 같은 과정으로 checkAndRegisterChkSum 단계까지 실행한 후에 백업할 위치를 찾아서 R로 반환하고, 복제후 복제된 파일의 체크섬을 C3

으로 반환하여 C2값과 C3값을 비교 후 저장한다.

그림 11은 기록물철의 생성을 위한 유저 인터페이스이다. 기록물철의 생성 시에 해당 기록물철의 데이터를 iCAT DB에 정의한 기록물 철 테이블에 입력하게 된다. 등록된 기록물 철에 대한 참조 값을 가지고 기록물건을 등록하게 된다. 기록물건은 기록물철의 참조 값과 기록물건의 기록유형을 참조 값으로 가지고 이 값을 기록물에 대한 사용자 정의 메타 데이터로써 iCAT DB에 저장되어 관리된다. 그림 12는 기록물건 등록의 유저 인터페이스이다. 기록물건 등록 시에 입력되는 데이터들은 iCAT DB에 사용자 정의 메타 데이터로 저장되며, 저장된 사용자 정의 메타 데이터는 검색 및 해당 기록물건의 정보 제공에 사용된다.

그림 13은 이러한 기록물의 등록시 메타 데이터와 실제 파일의 이동을 도식화한 것이다. 기록물이 웹상에 생

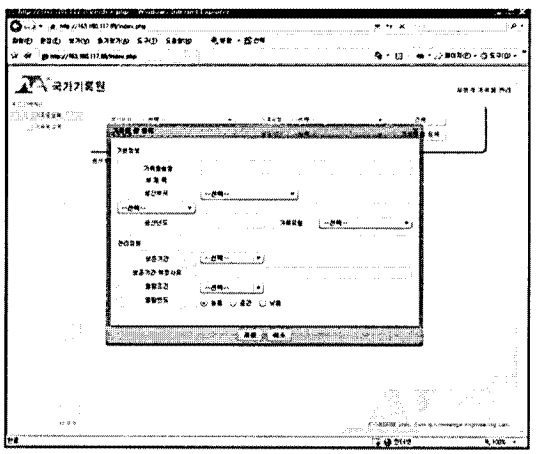

그림 11 기록물 철의 생성

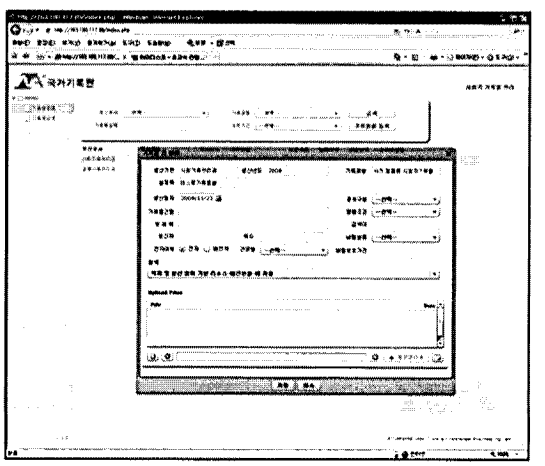

그림 12 기록물건의 등록

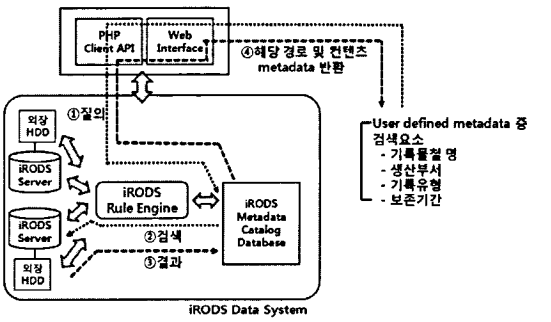

그림 13 기록물의 등록시 데이터의 흐름

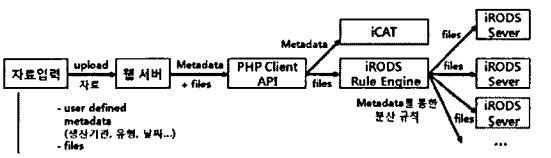

그림 14 기록물 검색 흐름

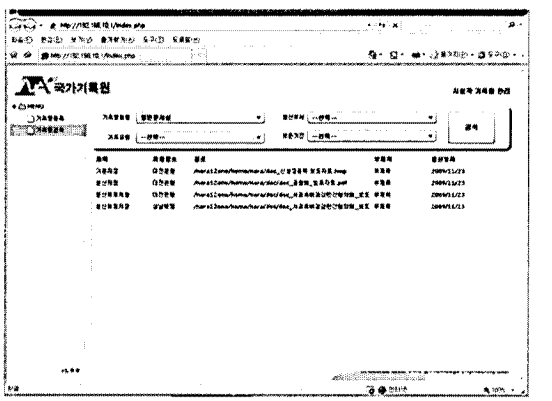

그림 15 메타데이터를 이용한 검색화면

성된 유저 인터페이스를 통하여 해당 기록물의 정보와 같이 웹서버로 전송되면, 웹서버는 메타데이터와 파일을 php 클라이언트 API를 통하여 각각 iCAT DB와 iRODS Rule 엔진에 전달하여 제공된 메타데이터에 따라 실시간으로 룰을 적용하여 iRODS 서버에 파일을 저장하는 과정을 나타낸다. 저장 과정 중 데이터의 복제를 통한 분산은 원본의 체크섬 값과 분산되어 저장된 복제본의 체크섬 값을 비교하여 복제본의 정당성을 확인하고, 저장된 기록물을 검색하여 호출할 때, 원본과 복제본의 체크섬 값을 다시 비교하여, 데이터의 무결성을 보장한다.

이러한 저장과정을 거친 기록물은 그림 14에 나타내어진 iCAT DB의 검색을 통하여 분산된 기록물의 위치와 메타데이터 정보를 취득함으로써 기록물을 관리할 수 있도록 구현하였다.

그림 15는 검색자가 검색조건을 선택하여, iRODS에서 제공하는 php API를 이용하여 메타데이터를 검색한 결과를 반환한 화면이다. 요구사항의 가정에 준하는 분산 및 복제결과를 적절히 반환함을 확인하였다.

6. 결 론

본 논문에서는 전자기록물 관리시스템의 설계 및 구현에 대해 논의하였다. 이러한 과정에서 규칙 기반 데이터 그리드 시스템인 iRODS를 이용한 설계 및 구현 이슈들을 논의하고 해결 방법을 제시하였으며, 기록물의 저장, 분산, 복제의 프로세스의 자동화와 저장 자원들을 논리적으로 가상화함으로써 시스템의 확장성과 유연성을 향상할 수 있음을 보였다. 또한 iRODS의 설정 값과 슬라이딩 윈도우의 사이즈를 조절함으로써 대용량 데이터의 전송속도를 최적화하여 효율적인 전송을 위한 방법을 제시하였다.

향후 전자기록물 관리 표준인 OAIS에서 요구하는 모든 기록물 관리프로세스를 자동화하기 위한 iRODS 규칙의 설계 및 구현에 대한 심도 있는 연구가 요구된다.

참 고 문 헌

[1] E. Jung, "Developing Strategies for Reliable Transfer of Electronic Records," Master Thesis, Hankuk University of Foreign Studies, 2007.

[2] D. Yun, "Proposal on Electronic Records Management System - Focused on Electronic Record Concept and International Standards," *Archive Management and Preservation*, vol.10, pp.1-32, 2005.

[3] S. Oh, I. Kim, J. Kim, "Technology Trends of Workflow Management for Automating Electronic Business Processes," *Trend Analysis of Electronic and Telecommunication*, vol.17, no.2, pp.61-70, April. 2002.

[4] S. Choi, J. Yang, J. Han, "Trend and Definition of BPM Technology Market," *Journal of Korea Information Processing Society*, vol.12, no.3, pp.18-29, May. 2005.

[5] R. W. Moore, A. Rajasekar, M. Wan, "Data Grids, Digital Libraries, and Persistent Archives: An Integrated Approach to Sharing, Publishing, and Archiving Data," *Special Issue of the Proceedings of the IEEE on Grid Computing*, vol.93, no.3, pp.578-588, March. 2005.

[6] iRODS Documentation, http://www.irods.org/index.php/Documentation, 2010.4.

[7] S. Lee, "Standardization of Digital Archiving and OAIS Reference Model," *Journal of Information Management*, vol.33, no.3, pp.45-68, 2002.

[8] CCSDS 650.0-B-1, "Reference Model for an Open Archival Information System(OAIS)," BLUE BOOK, Issue 1, 2002.

[9] B. Jacob, "Introduction to grid computing," IBM Red Book, 2005.12.

[10] K. Krauter, "A Taxonomy and Survey of Grid Resource Management Systems for Distributed Computing," *Journal of Software-Practice Experience*, pp.135-164, 2002.

[11] Using iRODS Data System Archive, http://www.diceresearch.org/DICD_Site/iRODS_Use_Cases/Archive.html, 2010.4.

[12] National Archives of Korea, "The Guidelines for Digitization [Scanning, Encoding]," Technical Documents, 2004.

[13] Sashka Davis, "Progress towards Efficient Data Ingestion into iRODS," Technical Report, 2008.9.

[14] Reagan W. Moore, "iRODS:integrated Rule Oriented Data System," White Paper, 2008.9.

한 용 구

2005년 경희대학교 컴퓨터공학과 학사
2007년 경희대학교 컴퓨터공학과 석사
2007년~현재 경희대학교 대학원 컴퓨터공학과 박사과정. 관심분야는 행위인지, 클라우드 컴퓨팅, 데이터 마이닝

김 진 승

2005년 경희대학교 물리학과 학사. 컴퓨터공학과 학사. 2007년 경희대학교 컴퓨터공학과 석사. 2009년~현재 경희대학교 대학원 컴퓨터공학과 박사과정. 관심분야는 분산 컴퓨팅, 데이터 마이닝

이 승 현

2009년 나사렛대학교 정보통신과 학사
2009년~현재 경희대학교 대학원 컴퓨터공학과 석사과정. 관심분야는 전자 기록물 관리, 경량 행위인지

이 영 구

1992년 한국과학기술원 과학기술대 전산학과 학사. 1994년 한국과학기술원 전산학과 석사. 2002년 한국과학기술원 전산학과 박사. 2002년 9월~2004년 2월 미국 UIUC 전산학과 Post Doctoral Research Fellow, 2004년 3월~현재 경희대학교 컴퓨터공학과 부교수. 관심분야는 대용량 데이터 관리, 클라우드 컴퓨팅, 유비쿼터스 데이터관리, 데이터 마이닝