

Lustre 병렬파일시스템 오토 프로비저닝을 위한 Ambari 서비스 설계에 관한 연구

곽재혁*, 김상완*, 변은규*, 남덕윤*
 *한국과학기술정보연구원
 e-mail:jhkwak@kisti.re.kr

A Study on the Design of Ambari Service for Lustre Parallel File System Auto Provisioning

Jae-Hyuck Kwak*, Sangwan Kim*, Eunkyu Byun*, Dukyun Nam*
 *Korea Institute of Science and Technology Information

요 약

하둡은 대표적인 빅데이터 처리 프레임워크로 널리 사용되고 있지만 하둡 어플리케이션은 고성능 컴퓨팅 환경에서 하둡 분산파일시스템이 아닌 러스터 병렬 파일시스템 위에서도 수행될 수 있다. 그러나 이를 위해서 추가적으로 러스터 병렬파일시스템을 구축하고 관리하는 것은 시간 소모적인 업무가 될 수 있다.

본 연구는 러스터 병렬파일시스템의 오토 프로비저닝을 위한 암바리 서비스의 설계 방안에 대해서 제안한다. 암바리는 하둡 클러스터의 프로비저닝, 관리, 모니터링을 위한 운영 관리 프레임워크이며 운영자의 필요에 따라서 확장할 수 있는 서비스 프레임워크를 제공한다. 본 연구에서는 암바리를 통해서 러스터 병렬파일시스템을 오토 프로비저닝하고 관리하기 위한 확장 서비스를 설계하였으며 서비스를 위한 컴포넌트와 각 컴포넌트별 중요한 기능 사항에 대해서 논하였다.

1. 서론

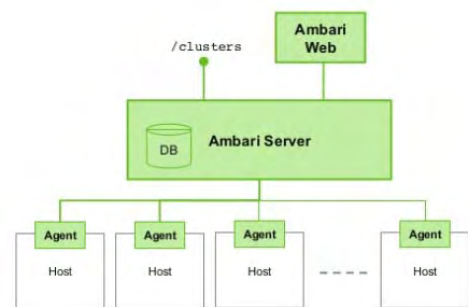
빅데이터 분석과 고성능 컴퓨팅은 별개의 영역으로 진행되어 왔지만 최근에는 고성능컴퓨팅 기반의 빅데이터 분석이 주목받고 있는 등 영역간의 융합이 가시화되고 있다 [1]. 예를 들어, 대기과학 시뮬레이션에 각 지역에서 수집되는 기후관련 데이터를 접합하여 더욱 정확한 예측이 가능할 수 있다.

하둡[2][3]은 대표적인 빅데이터 처리 프레임워크로 널리 사용되고 있지만 하둡 작업은 고성능컴퓨팅 환경에서 하둡 분산파일시스템이 아닌 러스터 병렬파일시스템[4][5] 위에서도 수행될 수 있다. 그러나 이를 위해서 추가적으로 러스터 병렬파일시스템을 프로비저닝하고 관리하는 것은 시간 소모적인 업무가 될 수 있다.

본 연구는 러스터 병렬파일시스템의 오토프로비저닝을 위한 운영관리 서비스의 설계 방안에 대해서 제안한다. 본 연구에서는 하둡 클러스터의 프로비저닝, 관리 모니터링을 위한 오픈소스 운영 관리 프레임워크인 암바리[6]를 사용하였으며 암바리에서 제공하는 확장 가능한 서비스 프레임워크를 기반으로 러스터 병렬파일시스템을 오토프로비저닝하고 관리하기 위한 LustreMgmtService를 설계하였고 서비스를 위한 컴포넌트와 각 컴포넌트별 중요한 기능 사항에 대해서 논하였다.

2. 암바리 및 러스터 소개

암바리는 하둡 클러스터의 프로비저닝, 관리, 모니터링을 위한 소프트웨어를 개발함으로써 하둡 관리를 단순하게 하기 위한 목적을 가지고 시작된 아파치 오픈소스 프로젝트이다. 클러스터에 하둡 서비스를 설치하고 설정 변경을 할 수 있으며 하둡 서비스를 시작, 종료 재구성할 수 있는 중앙 관리가 가능하고 하둡 클러스터의 상태를 모니터링하기 위한 대쉬보드를 제공한다. 암바리는 REST API를 제공함으로써 어플리케이션에 암바리를 통합하는 것을 가능하게 한다. 암바리는 관리 노드에 설치되는 Ambari Server, 클러스터 노드에 설치되는 Agent, 웹 노드에 설치되는 Ambari Web로 구성되며 암바리의 시스템 아키텍처는 다음과 같다.



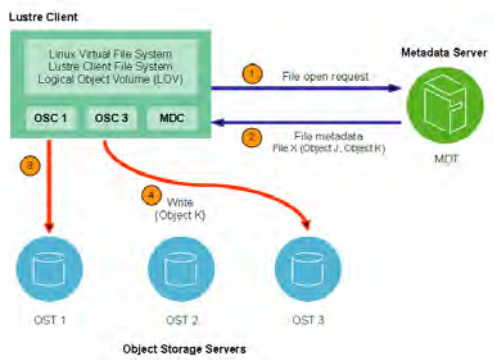
<그림 1> 암바리 시스템 아키텍처

암바리는 확장 가능한 서비스 프레임워크를 제공함으로써 커스텀 서비스를 개발하는 것을 가능하게 한다. 하나의 서비스는 다수의 컴포넌트로 구성되며 각 컴포넌트 카테고리별로 디폴트 라이프사이클 커맨드를 가지며 라이프사이클 커맨드도 확장이 가능하다. 예를 들면, 암바리의 YARN 서비스를 보면 ResourceManager는 마스터로, NodeManager는 슬레이브로, YarnClient는 클라이언트로 정의되어 있으며 NodeManager는 decommission이라는 커스텀 커맨드를 가지고 있음을 확인할 수 있다.

<표 1> 암바리 서비스의 컴포넌트 카테고리 및 디폴트 라이프사이클 커맨드

Component Category	Default Lifecycle Commands
MASTER	install, start, stop, configure, status
SLAVE	install, start, stop, configure, status
CLIENT	install, configure, status

러스터는 슈퍼컴퓨터와 같은 고성능 시스템에서 널리 사용되는 병렬파일시스템으로서 하둡 분산파일시스템과는 달리 포직스 표준을 따르는 파일시스템 인터페이스를 제공하고 있어서 고성능의 공유파일시스템으로서 사용된다. 러스터는 메타데이터 정보를 관리하는 MDS/MDT, 객체 기반의 데이터를 저장하는 OSS/OST로 구성되어 있다. 러스터에서 파일 데이터는 스트라이프(stripe)라는 객체 단위로 쪼개져서 복수개의 OST에 분산되어 저장되며 클라이언트는 OST와 직접적으로 통신하여 병렬 입출력을 통한 성능을 보장받는다. 러스터는 고속의 인피니밴드 네트워크에 최적화되어 있어서 파일시스템 성능을 높이는 요소로 활용된다.



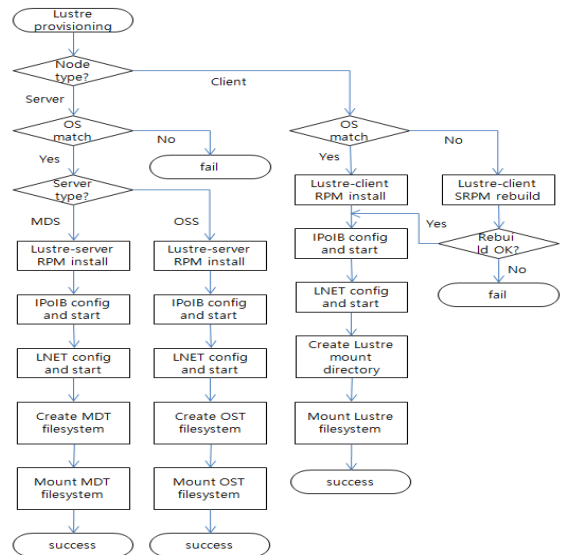
<그림 2> 러스터 병렬파일시스템 구조

3. LustreMgmtService 설계

본 논문에서는 러스터 병렬파일시스템을 오토프로비저닝하고 관리하기 위해 LustreMgmtService라는 암바리 서비스를 정의하였다. LustreMgmtService는 마스터로 구분된 4개의 컴포넌트와 클라이언트로 구분된 1개의 컴포넌트를 가지며 러스터 오토 프로비저닝 및 관리 기능 이외에 Hadoop-on-Lustre 어댑터[7][8]와 하둡 클러스터와 러스터 병

렬파일시스템 노드 간의 계정 동기화 기능도 포함하고 있다.

러스터 병렬파일시스템의 프로비저닝 순서도는 다음과 같이 정의하였다. 러스터 레포지토리에 러스터가 지원하는 운영체제 버전에 맞게 서버와 클라이언트 RPM이 제공한다. 러스터가 지원하는 운영체제가 아닌 경우라면 SRPM을 리빌드하여 운영체제에 맞는 RPM을 생성할 수 있지만 서버 RPM의 경우는 커널을 바꾸기 때문에 리빌드하더라도 실패할 가능성이 높다. 다만, 클라이언트 RPM의 경우 커널을 바꾸지 않고 모듈로만 설치되므로 리빌드하는 경우에 성공할 가능성이 높아서 이를 순서도에 반영하였다.



<그림 3> 러스터 병렬파일시스템 프로비저닝 순서도

LustreMgmtService의 각각의 컴포넌트에 대한 기능 및 커맨드별 역할은 다음과 같다.

○ HadoopLustreAdapterMgmtService (MASTER)

하둡에서 HDFS를 사용하지 않고 러스터를 사용할 수 있게 하는 Hadoop-on-Lustre 어댑터를 위한 관리 데몬으로서 동작한다. 라이프사이클 커맨드의 역할은 다음과 같이 정의하였다.

<표 2> HadoopLustreAdapterMgmtService 커맨드 및 역할

커맨드	역할
install	HadoopLustreAdapterMgmtService 관리 데몬을 설치
configure	Hadoop-on-Lustre 어댑터의 설치하고 switch-to-lustre 커맨드를 호출하여 Hadoop-on-Lustre 기능을 사용하기 위한 Hadoop Configuration Parameter 변경
start	HadoopLustreAdapterMgmtService 관리 데몬 시작
stop	HadoopLustreAdapterMgmtService 관리 데몬 종료
switch-to-lustre	Hadoop-on-Lustre를 위한 Hadoop Configuration Parameter 변경
switch-to-hdfs	디폴트 Hadoop을 위한 Hadoop Configuration Parameter 변경

○ LustreMDSMgmtService (MASTER)

러스터 병렬파일시스템의 메타데이터 관리 서버인 MDS를 관리하기 위한 관리 데몬으로서 동작한다. 라이프사이클 커맨드의 역할은 다음과 같이 정의하였다.

<표 3> LustreMDSMgmtService 커맨드 및 역할

커맨드	역할
install	LustreMDSMgmtService 관리 데몬을 MDS노드에 설치, Lustre 레포지토리로부터 Lustre Server RPM 다운로드 및 설치
configure	IPoIB 네트워크 설정 및 시작, LNET 설정 및 시작, MDT 파일시스템 생성 및 마운트
start	LustreMDSMgmtService 관리 데몬 시작
stop	LustreMDSMgmtService 관리 데몬 종료
mount-mdts	MDT 파일시스템 마운트
unmount-mdts	MDT 파일시스템 언마운트
mkfs-mdts	MDT 파일시스템 생성

○ LustreOSSMgmtService (MASTER)

러스터 병렬파일시스템의 오브젝트 저장 서버인 OSS를 관리하기 위한 관리 데몬으로서 동작한다. 라이프사이클 커맨드의 역할은 다음과 같이 정의하였다.

<표 4> LustreOSSMgmtService 커맨드 및 역할

커맨드	역할
install	LustreOSSMgmtService 관리 데몬을 OSS노드에 설치, Lustre 레포지토리로부터 Lustre Server RPM 다운로드 및 설치
configure	IPoIB 네트워크 설정 및 시작, LNET 설정 및 시작, OST 파일시스템 생성 및 마운트
start	LustreOSSMgmtService 관리 데몬 시작
stop	LustreOSSMgmtService 관리 데몬 종료
mount-osts	OST 파일시스템 마운트
unmount-osts	OST 파일시스템 언마운트
mkfs-osts	OST 파일시스템 생성

○ UserSyncService (MASTER)

OpenLDAP을 이용하여 하둡 클러스터와 러스터 병렬파일시스템 노드 간의 계정 정보 동기화를 수행한다. 라이프사이클 커맨드의 역할은 다음과 같이 정의하였다.

<표 5> UserSyncService 커맨드 및 역할

커맨드	역할
install	OpenLDAP RPM 다운로드 및 설치
configure	OpenLDAP 설정 파일 생성
start	OpenLDAP 서버 데몬 시작, 최초 시작일 경우 OpenLDAP 관리자 비밀번호 생성, 기본 스키마 로드, 도메인 이름 설정
stop	OpenLDAP 서버 데몬 종료

○ LustreClient (CLIENT)

러스터 병렬파일시스템의 클라이언트와 OpenLDAP 클라

이언트를 설치하고 설정한다. 라이프사이클 커맨드의 역할은 다음과 같이 정의하였다.

<표 6> LustreClient 커맨드 및 역할

커맨드	역할
install	Lustre 레포지토리로부터 Lustre Client RPM 다운로드 및 설치, OpenLDAP Client RPM 다운로드 및 설치
configure	러스터 마운트 디렉토리 생성 및 마운트, OpenLDAP 구성
mount-lustre	하둡 노드의 러스터 마운트
unmount-lustre	하둡 노드의 러스터 언마운트

4. 결론 및 향후 계획

본 연구는 하둡 클러스터를 위한 오픈소스 운영 관리 프레임워크인 암바리 기반으로 러스터 병렬파일시스템의 오토프로비저닝을 지원하는 운영관리 서비스를 설계하고 서비스 컴포넌트 및 중요한 기능 사항에 대하여 논하였다. 본 연구에서 제안하고 있는 LustreMgmtService는 현재 구현 중에 있으며 이 서비스가 구현된다면 빅데이터 운영 관리 프레임워크인 암바리를 사용하여 러스터 병렬파일시스템에 대한 운영 관리도 가능하며 하둡 분산파일시스템이 아닌 러스터 병렬파일시스템 기반으로 하둡 어플리케이션을 수행하는 것이 가능할 것으로 생각된다.

※ 이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원(No. R0190-17-2012, 빅데이터 처리 고도화 핵심기술개발 사업 총괄 및 고성능 컴퓨팅 기술을 활용한 성능 가속화 기술 개발)과 한국과학기술정보연구원 주요사업의 지원을 받아 수행된 연구임

참고문헌

[1] S. Conway, C. DeKate, "High-Performance Data Analysis: Big Data Meets HPC", IDC Directions Conference 2013, 2013.
 [2] Apache Hadoop, <http://hadoop.apache.org>
 [3] T. White, Hadoop: The Definitive Guide, OREILLY, 2010.
 [4] Lustre, <http://lustre.org>
 [5] Lustre Operations Manual, http://doc.lustre.org/lustre_manual.pdf
 [6] Ambari, <https://cwiki.apache.org/confluence/display/AMBAR/Ambari>
 [7] O. Kulkarni, "Hadoop MapReduce over Lustre", Lustre User Group Conference 2013, 2013.
 [8] 광재혁, 김상완, 허태상, 황순욱, "러스터 파일 시스템 기반 하둡 맵리듀스 실행 환경 구현 및 성능 분석", 정보과학회 컴퓨팅의 실제 논문지 제21권 제8호, 2015.08.