

통합 로그 저장소를 이용한 슈퍼컴퓨터 상태 분석 도구 설계

김성준*, 이재국*,홍태영*

*한국과학기술정보연구원 슈퍼컴퓨팅인프라실

e-mail:sjkim@kisti.re.kr

Design of tool for supercomputer status analysis using log integrated repository

Sung-Jun Kim*, Jae-Kook Lee*, Tae-Young Hong*

*Dept of Supercomputing Infra, KISTI

요 약

슈퍼컴퓨터와 같은 대규모 클러스터는 계산노드, 파일시스템, 인터컨넥트 네트워크 등과 같은 다양한 요소로 구성된다. 연구자들은 자신의 작업을 계산노드들을 이용하여 병렬화된 계산 작업을 수행한다. 수천 노드로 구성된 슈퍼컴퓨터에서 연구자들의 계산 작업이 효율적으로 수행되는 것을 파악하는 것은 시스템 관리자들에게 어려운 문제이다. 본 논문에서는 슈퍼컴퓨터에서 생성되는 다양한 로그를 통합 저장하고 이를 기반으로 슈퍼컴퓨터의 활용 효율을 제고하기 위한 통합 로그 저장소를 설계하고, 향후 구현될 슈퍼컴퓨터 상태 분석 시스템에서 활용할 수 있는 몇 가지 시나리오를 통해서 본 논문에서 제안하는 통합 로그 저장소의 효용성을 설명한다.

1. 서론

한국과학기술정보연구원의 슈퍼컴퓨팅서비스센터에서는 초대형 슈퍼컴퓨터를 국내 연구자들의 연구에 활용할 수 있도록 제공하고 있다. 연구자들은 과학계산을 위해서 병렬 프로그래밍을 작성하고, 이를 슈퍼컴퓨터상의 작업관리 프로그램(이하 스케줄러)에 전달하면, 스케줄러는 유휴 계산자원에 이를 할당하고 계산을 수행하는 방식으로 운영된다.

연구자들의 프로그램 특성 혹은 완성도에 따라서 계산자원을 효율적으로 활용하지 못하는 경우가 간혹 발생한다. 이 경우 연구자들은 수행시간의 증대로 인하여 원하는 시간내에 연구 결과를 받아 볼수 없으며, 서비스센터의 입장에서는 자원의 비효율적인 운영으로 인해서 보다 많은 연구자들에게 서비스를 제공해야 하는 기회 감소로 이어지게 된다.

현재 서비스센터에서는 ganglia와 같은 클러스터 모니터링 프로그램 및 자체 개발한 다양한 도구를 활용하여 시스템의 활용도를 모니터링하고 통계 데이터를 산출하여 운영 효율성을 주기적으로 관리하고 있다.

하지만, 기존 모니터링 도구들은 연구자들의 작업과 시스템의 이용률을 연계하여 확인할 수 있는 방법이 없어, 시스템 관리자들은 각각의 도구에서 생성된 정보를 각각 확인하여 시스템의 현황을 파악해야 하는 단점이 있었다.

본 논문에서는 사용자 작업 정보(수행시작시간, 종료시간, 할당된 노드 정보 등)를 스케줄러를 통해서 수집하고,

이를 Ganglia와 같은 모니터링 도구에서 추출한 시스템의 이용률과 연계하여 전체적인 시스템의 효율성을 연구자들의 작업을 기준으로 쉽게 파악할 수 있도록 통합 로그 저장소를 설계하고자 한다. 본 고에서 제안된 저장소의 구성은 향후 슈퍼컴퓨터 상태 분석 도구의 구현에 활용될 예정이다.

2. 관련 연구

2.1 SGE(Sun Grid Engine)

다음의 <표 1>는 SGE가 생성하는 로그 중에 과금 관련된 일부 정보만을 표현한 것이다. 로그에는 작업을 수행한 사용자의 아이디, 작업이 사용한 노드정보, 작업이 노드를 사용한 기간 등의 정보가 포함되어 있다.

<표 1> SGE 과금 관련 정보

| | | |
|-------------|---------------|------------|
| queue | wait time | cpu time |
| exec_host | run time | job option |
| group | mem usage | exit code |
| userid | vmem usage | pe_taskid |
| job id | number of cpu | job name |
| submit time | start time | end time |

2.2 Ganglia

Ganglia는 대표적인 클러스터 모니터링 도구이다. 이 도구는 개별 노드의 성능정보를 모니터링 하는 gmond와 이

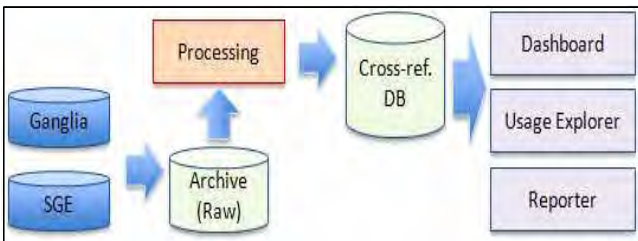
를 정보를 취합하여 저장하는 gmetad로 구성된다. 다음의 <표 2>은 gmond에서 모니터링 하는 노드의 성능 정보 중 일부를 표현한 것이다.

<표 2> Ganglia 모니터링 Metrics

| 구분 | 모니터링 정보 |
|---------|-----------------------------------|
| CPU | Idle, User, System, wio |
| DISK | Space available, Total disk space |
| Load | Load {one five fifteen } |
| Memory | Buffers, Cached, Free, Shared |
| Network | {Byte Packet} {Received Sent} |

3. 설계

앞서 살펴본 바와 같이, SGE는 연구자의 작업과 관련된 정보를 포함하고 있으며, Ganglia는 노드의 상태에 대한 정보를 포함하고 있다. 하지만, 시스템의 효율적인 운영을 위해서는 이들 간의 정보를 종합적으로 파악하고 대응을 해야 하는 필요가 있다. 이를 위해서 다음의 <그림 1>과 같이 슈퍼컴퓨터 상태 분석을 위한 통합 로그 저장소를 설계하고 로그들의 관계를 분석하여 저장하고 사용 현황 파악 및 통계 산출을 위해서 활용할 수 있을 것이다.



(그림 1) 시스템 구조도

상기 시스템을 이용하여 다음과 같이 기존 개발 모니터링 도구에서는 파악하기 어려운 정보를 확인할 수 있을 것이다.

<표 3> 시나리오 1

| | |
|-------|---|
| 상황 1 | 특정 연구자의 작업이 사용한 계산노드에서의 실제 CPU 활용 상태 |
| 참조 정보 | SGE : 연구자ID, 작업할당 노드정보, 작업시작시간, 작업종료시간 Ganglia: 측정시간, CPU 부하율 정보 |
| 예상 화면 | |

상기 <표 3>은 본 논문에서 설계한 통합 로그 저장소를 이용하여 구성 가능한 화면 중에 하나를 예시로 표현하였다. 기존 개별 모니터링 도구를 이용해서는 파악하기 어려운 정보를 화면을 통해서 표출이 가능할 것으로 예상 이 된다.

예상화면에서는 특정 연구자의 작업이 계산노드들을 제대로 활용하지 못하는 문제를 파악하기 용이하며, 이를 파악하여 연구자의 작업 코드를 수정함으로써 작업 효율 향상 및 시스템 활용률 제고에도 활용할 수 있을 것이다.

이외에도 연구자의 작업과 파일시스템의 부하율 혹은 네트워크 부하율 등과의 관계도 차트 형태로 표출하여 연구자 작업이 시스템에 미치는 영향관계 파악에도 도움이 될 것이다.

로그를 저장할 데이터베이스는 기존의 관계형 데이터베이스가 아닌 비정형 데이터 저장에 적합한 MongoDB와 같은 NoSQL 데이터베이스를 이용할 것이다. 이는 생성되는 로그의 양과 정형화를 했을 때 발생하는 데이터의 낭비를 최소화하고 보다 용이한 화면 표출을 위해서 데이터의 스키마를 정의하기 위해서이다.

또한, 보고서 출력에 사용할 정보는 배치작업의 형태로 일일단위로 결산한 결과를 미리 계산하여 저장 함으로써, 보고서 생성시 소요되는 시간을 최소화 하여 보고서 생성을 위한 쿼리 소요 시간을 최소화 할 것이다.

4. 결론

본 논문에서는 제안하는 슈퍼컴퓨터 상태 분석 도구는 슈퍼컴퓨터 활용과 관련되는 작업관리프로그램(SGE)와 성능 모니터링 도구(Ganglia)의 로그를 통합 저장하고, 이를 조합함으로써 기존 단일 모니터링 도구에서는 파악하기 힘든 정보를 생성하는 것을 목표로 한다.

향후 본 설계를 바탕으로 실제 시스템을 구축하면 시스템 관리자는 보다 쉽게 시스템에 대한 활용현황을 모니터링 및 분석 할 수 있을 것이다. 또한, 이를 활용하여 연구자의 작업 효율 향상 및 시스템의 활용률 제고하는데 도움이 될 것으로 예상된다.

참고문헌

[1] Ganglia Monitoring System, <http://ganglia.info>
 [2] Sun Grid Engine Accountfile Format, <http://grid-scheduler.sourceforge.net/htmlman/htmlman5/accounting.html>
 [3] SungJun Kim, et al, "Analysis of user batchc job statistics to maximize HPC resource usage", ICSMB, pp.323-324,2016
 [4] SungJun Kim, et al, "User Job pattern analysis of KISTI 4th SMP supercomputer", ICCT2016, pp.1076-1077, 2016