대규모 클러스터 시스템에서 전력 사용량 측정 및 분석 방법에 관한 연구

이재국[1], 안도식[1], 홍태영[2]
[1]한국과학기술정보연구원 슈퍼컴퓨팅인프라센터
[2]한국과학기술정보연구원 슈퍼컴퓨팅인프라센터장
jklee@kisti.re.kr, dsan@kisti.re.kr, tyhong@kisti.re.kr

Measurement and Analysis of Power Consumption in a Large-scale Cluster System

Jae-Kook Lee[1], Dosik An[1], Taeyong Hong[2]
[1]Div. of Supercomputing Infrastrucutre, KISTI
[2]Derector, Div. of Supercomputing Infrastrucutre, KISTI

요 약

슈퍼컴퓨터와 같은 대규모 클러스터 시스템은 연산 성능이 높아지면서 소비 전력도 증가하고 있다. 본 논문에서는 블레이드 형태의 대규모 클러스터 시스템에서 응용 프로그램이 수행될 때 소비되는 전력량을 측정하는 방법을 제안하고 실제 운영중인 국가 슈퍼컴퓨터 5호기 시스템에서 수집된 230만 개 이상의 응용 프로그램 수행 데이터와 전력 소비량 데이터를 이용하여 분석한다.

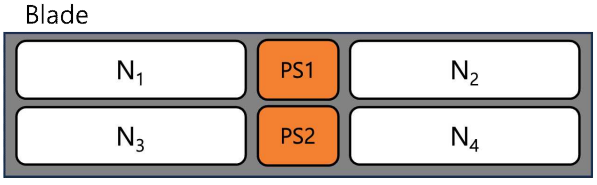
1. 서론

슈퍼컴퓨터와 같은 대규모 클러스터 시스템은 계산 수요의 증가와 함께 많은 소비전력을 요구한다. 슈퍼컴퓨터 TOP500의 1위인 'El Capitan' 슈퍼컴퓨터는 이론성능(Rpeak) 2.75EFlops이며 약 30MW의 전력을 소비한다. 2위 시스템인 'Frontier'는 Rpeak 2.06EFlops에 약 24.6MW, 3위 시스템인 'Aurora'도 Rpeak 1.98EFlops에 약 38.7MW의 전력을 소비한다[1]. 이처럼 연산 성능의 증가에 따른 전력 소비량의 증가 문제로[2]. 에너지 효율성을 높이기 위한 연구들이 지속되고 있다[3]. 국가 슈퍼컴퓨터 5호기 누리온 뿐만 아니라 차기 슈퍼컴퓨터의 에너지 효율성 개선을 위해서는 먼저 전력 소비량의 측정이 선행되어야 한다. 대규모 클러스터 시스템의 경우 집적도를 높이기 위해서 전력공급장치(PSU)나 냉각팬 등을 공유하는 블레이드 형태로 구성된다. 국가 슈퍼컴퓨터 5호기 누리온도 8,437대 계산노드로 구성된 초거대 클러스터 시스템으로 블레이드 형태의 시스템이다[4]. 누리온은 나이츠랜딩(KNL)과 스카이레이크(SKL)라는 코드명을 가진 2가지 타입의 CPU로 구성이 되어있다. KNL 계산노드는 노드당 68코어(1.4GHz)를 갖는 8,305개로 구성되어 있고, SKL 계산노드는 40코어(2.4GHz)를 갖는 132개 노드로 구성되어 있다. 모든 계산노드들과 데이터가 저장되는 병렬파일시스템은 OPA라는 초고속 인터커넥트로 모두 연결되어 있다[5].

본 논문에서는 블레이드 형태의 누리온 시스템에서 각 계산 노드의 전력 데이터를 측정하는 방법 제안하고 누리온에서 수집된 전력 소비량 데이터를 이용하여 전력 소비량에 영향을 미치는 요소를 분석한다.

2. 블레이드 서버

(그림 1)은 누리온 시스템의 2U 크기 인클로저(Enclosure)를 간략히 표현한 것이다. 하나의 인클로저에는 4개의 계산노드가 장착되어 있으며 이중화를 위해 2개의 PSU로 구성되어 있고 4개의 계산노드가 PSU를 공유하는 형태이다. 누리온의 42U 크기 랙(Rack)에는 18개의 인클로저가 장착되어 있어 랙당 72개의 계산노드가 정착되어 있는 형태이다. 랙에도 2개의 전력분배장치(PDU)를 통해 이중화되어 안정적인 시스템 운영이 가능하다.



(그림 1) 슈퍼컴퓨터 5호기 누리온 인클로저 형태

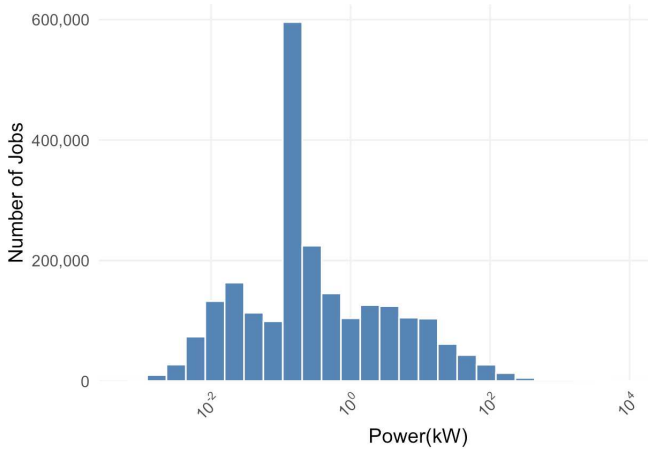
3. 전력 소비량 측정 및 분석

누리온은 2개의 PSU를 4개의 계산노드가 공유해서 사용하는 블레이드 형태이므로 각 계산노드의 전력 소비량을 측정하기 위해서 CPU 사용률을 가중치로 사용한다. 임의의 시점 (t)에서 계산노드의 CPU 사용률을 $u_j(t)$, $j=1,2,3,4$ $(u_j \geq 0)$라 할 때, 블레이드 전체 CPU 사용률은 $S(t)=\sum_{k=1}^{4} u_j(t)$이다. (t) 시점에 각 계산노드의 CPU 사용율 가중치$(w_j(t))$는 다음과 같다.

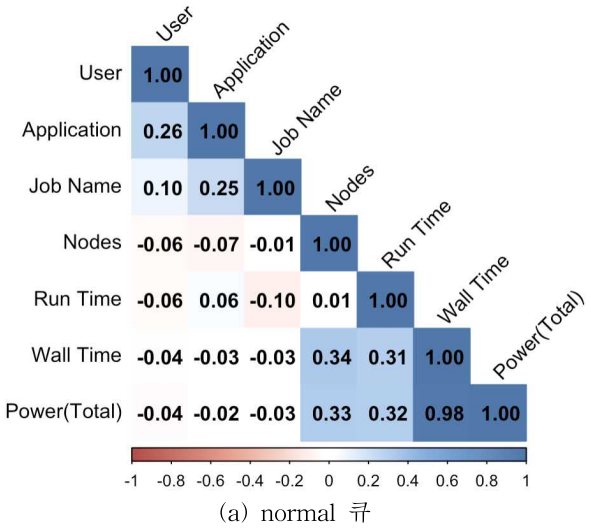
$$w_j(t)=\begin{cases}\dfrac{u_j(t)}{S(t)}, & S(t)>0\\ \dfrac{1}{4}, & S(t)=0\end{cases}$$

2개 PSU의 전략 사용량을 각 각 $P_1(t), P_2(t)$라 할 때, 블레이드의 총 전력 소비량은 $T(t)=P_1(t)+P_2(t)$이고, CPU 사용율 가중치를 적용하면 임의의 시점 (t)의 계산노드별 전력 소비량은 $P_j(t)=T(t)w_j(t)$이 된다. 각 계산노드에서는 1분 단위로 지정된 파일에 CPU 사용율과 전력 소비량을 측정하여 지정된 파일에 기록한다.

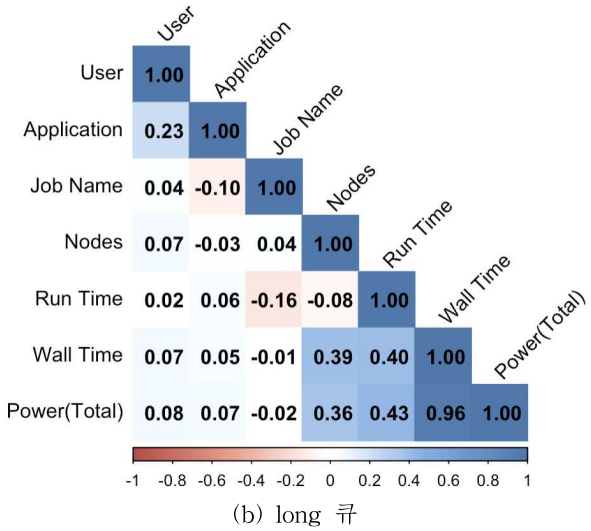
본 논문에서는 2024년도 1월 1일부터 12월 31일까지 누리온에서 실행된 약 2,302,375개 작업에 대한전력 소비량 데이터를 분석한다. (그림 2)는 전력 소비량에 따른 작업 개수를 막대 그래프로 나타낸 것이다. 그래프에서 보면 전력 소비량 값이 로그 스케일로 오른쪽 꼬리가 긴 분포(Long-tailed Distribution)를 보이고 있다.



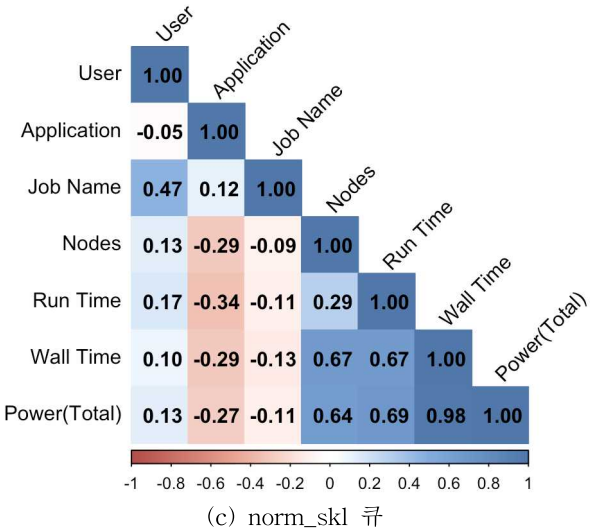
(그림 2) 전력 사용량에 따른 작업 분포



(a) normal 큐



(b) long 큐

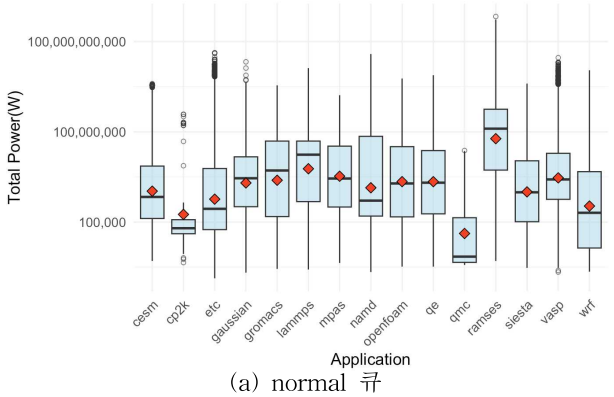


(c) norm_skl 큐

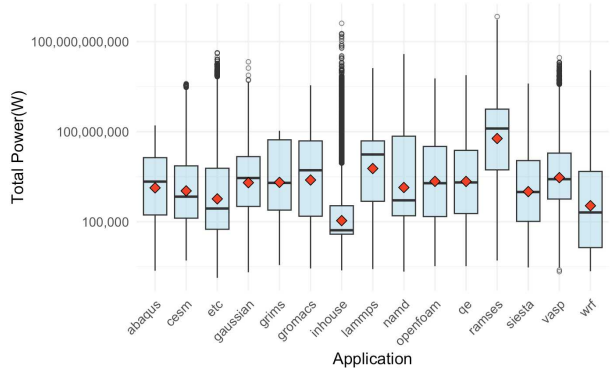
(그림 3) 큐별 데이터 속성 간 상관관계 분석

(그림 3)은 사용자(User), 응용 프로그램 종류(Application), 작업이름(Job Name), 노드개수(Nodes), 실행시간(Runing Time), 작업 수행시간(Wall Time), 전력소비량(Power) 데이터를 피어슨

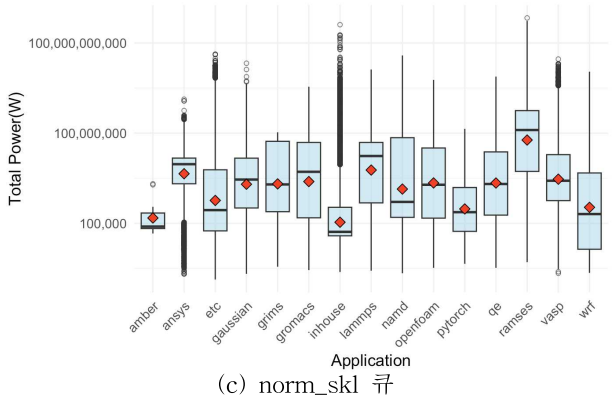
상관관계(Pearson Correlation) 분석 한 그래프이다. 누리온은 일반 사용자들에게 KNL 계산노드 사용을 위해 normal 큐를 제공하고, SKL 계산노드 사용을 위해 norm_skl 큐를 제공한다. 다만 이 큐들은 사용시간이 48시간으로 제한이 되어 장시간(120시간) 실행이 필요한 KNL 계산노드 사용자에게 long 큐를 제공한다. 전력 소비량과 가장 연관성이 높은 피처(Feature)는 작업 수행시간이다. 작업 수행시간은 각 응용 프로그램이 사용한 CPU 코어(core)의 개수에 실행시간을 반영한 값이다. 코어를 많이 사용하고 실행시간이 긴 응용 프로그램일수록 전력을 많이 소비하는 것을 확인할 수 있다. (그림 3)(c)에서 norm_skl 큐 사용자이 작업이름과 상관관계가 있는 것을 확인할 수 있었다.



(a) normal 큐



(b) long 큐



(c) norm_skl 큐

(그림 4) 응용 프로그램별 전력 사용량 분석

(그림 4)는 누리온에서 수행되는 응용 프로그램별 전력 소비가 많은 15개의 응용 프로그램을 박스 플롯(Box plot)으로 나타낸 것이다. 그래프에서 보는 것과 같이 같은 응용 프로그램 내에서도 전력 소비 패턴이 달라지는 것을 확인할 수 있다. 그래프의 y축 전력 소비량 값이 로그 스케일이므로 동일한 응용 프로그램이라도 전력 소비량 차이가 확연하게 나타나는 것을 확인할 수 있다. 다만 norm_skl 큐에서 실행된 분자 역학 시뮬레이션 소프트웨어인 Amber의 경우 전력 소비량의 패턴이 크게 달라지지 않은 것을 확인할 수 있다.

4. 결론

본 논문에서는 블레이드 형태로 PSU를 공유하는 계산노드에서 응용 프로그램이 실행될 때 소비하는 전력량을 측정하는 방법을 제안하고 2024년도 수집된 데이터를 이용하여 응용 프로그램별 전력 사용량 및 응용 프로그램이 수행될 때의 피처값과 전력 소비량의 상관관계를 분석하였다.

Acknowledgement

본 연구는 2025년도 한국과학기술정보연구원(KISTI) 기본사업 과제(K25L2M2C2)의 지원을 받아 수행한 것입니다.

참고문헌

[1] TOP500, (https://top500.org).

[2] Enterprise Viewpoint, "Sustainable High Performance Computing,“ (https://enterpriseviewpoint.com/sustainable-high-performance-computing/), 2024.

[3] T. Miyazaki, I. Sato, and N. Shimizu, "Bayesian optimization of hpc systems for energy efficiency," in ISC'18, pp. 44-62, Springer, 2018.

[4] 권민우, 윤준원, 홍태영, “PBS 작업 스케줄러 Hook를 이용한 슈퍼컴퓨터 5호기 계산노드 자동 검증 기능 구현,” 한국정보처리학회 학술대회논문집, 대한민국, 2019, pp.101-102.

[5] Jae-Kook Lee, Min-Woo Kwon, Do-Sik An, Junweon Yoon, Taeyong Hong, Joon Wu, Sung-Jun Kim, and Guohua Li, “Improvements to Supercomputing Service Availability Based on Data Analysis,” Appl. Sci. 11(13), 2021.