

The Electrical Grid and Supercomputer Centers: An Investigative Analysis of Emerging Opportunities and Challenges

Fname Lname¹, Fname Lname² Fname Lname, Fname Lname, Fname Lname,
Fname M. Lname, and Fname Lname

¹ Institute Name, City State Zip, Country,
Name@email.adr,

WWW home page: <http://web/page.html>

² Another Institute, nstitute Department, address1,
address2, Country

Abstract. Some of the largest supercomputer centers in the United States are developing new relationships with their electricity providers. These relationships, similar to other commercial and industrial facilities, are driven by mutual interest to save energy costs and improve electricity grid reliability. Supercomputer centers are concerned about electricity price, quality, environmental impact and availability. Electricity providers are concerned about supercomputer center's impact on the electrical grid reliability, both for energy consumption, peak power and fluctuations in power. Supercomputer center power demand can be greater than 20 megawatts (MW) or more, theoretical peak power requirements greater than 40 MW and re-occurring intra-hour variability can exceed 8MW. As a consequence, there are some supercomputer centers whose electricity providers are asking for hourly forecasts of power demand, a day in advance. This paper explores today's relationships, potential partnerships and possible integration between supercomputer centers and their electricity providers and its value. It develops a model for possible integration between supercomputer centers and the electrical grid. It then explores the utility of this model based on feedback from a questionnaire of Top 100 List sized supercomputer centers in the United States.

1 Introduction and Background

Supercomputer centers with petascale systems for high-performance computing (HPC) are realizing the large impact they will be putting on their electricity service providers with peak power demands of 20MW and instantaneous power fluctuations of 8MW.

The Energy Efficiency HPC Working Group (EE HPC WG) has been investigating opportunities for large supercomputing sites to more closely integrate with their electricity service providers. This paper documents the results of this investigative activity.

01

1/11/14

by: sridutt.

02

1/11/14

by: sridutt.

03

1/11/14

by: sridutt.

Leveraging prior work on data center and grid integration opportunities done by Lawrence Berkeley National Laboratory's (LBNL) Demand Response Research Center (<http://drcc.lbl.gov/publications>), this paper takes as a starting point LBNL's model for integrating data centers and the electrical grid. The model describes programs that are used by the electricity service providers to integrate with their customers (such as demand response) and methods used to balance the grid supply and demand of electricity. It also describes strategies that data centers might employ for managing their electricity and power requirements. This paper tuned this model's data center strategies for supercomputer centers. As opposed to data centers, supercomputer centers have very high system utilization and are not likely to use virtualization as a strategy. Also, supercomputer applications are generally not easily portable between geographic locations for a variety of reasons; security, data-locality, system tuning. Therefore, although included in the model for data centers, both virtualization and geographic load shifting were eliminated as potential supercomputing center strategies.

The first section of this paper describes in greater detail the model for integrating supercomputer centers and the electrical grid. The second section is a review of prior work on HPC center strategies that might be deployed for managing electricity and power. In order to further understand today's relationships, potential partnerships and possible integration between HPC centers, their electricity providers and the grid, a questionnaire was deployed whose respondents were Top100 List class supercomputer centers in the United States. The third section of this paper describes the results of that questionnaire. The fourth section of the paper describes opportunities, solutions and barriers. A fifth section describes conclusions and next steps. Finally, the last section recognizes additional authors.

2 Supercomputing Centers and Electrical Grid Integration

The EE HPC WG Team took as their starting point a model developed by LBNL's Demand Reponse Research Center ³ that describes ways in which data centers and electricity service providers may interact and how this integration can advance new market opportunities. This integration model describes programs that are used by the electricity service providers to encourage particular behaviors by their customers and methods used to balance the grid supply and demand of electricity. It also describes strategies that data centers might employ for utility programs to manage their electricity and power requirements and lower costs. The EE HPC WG Team adopted this model with slight tweaks to reflect the HPC environment (versus the general data center).

³ LBNL Data Center Grid Integration Activities: <http://drcc.lbl.gov/projects/dc>

2.1 Electricity Demand and Supply Side Management

Electricity providers use demand-side management for both energy efficiency and to balance the electricity supply. Demand side management programs are used to manage the consumption of energy on the consumers side of the meter. Demand side management can include energy conservation, energy efficiency and peak load management, and responses to changing supply conditions. The focus for this paper is on programs that are targeted at load management. Supply side management methods are used to ensure the efficient generation, transmission and distribution of electricity. These methods manage the supply of energy on the electricity providers side of the meter to meet the changing demand conditions.

There are many different ways that the end-users of electricity can modify or change their electricity consumption. The following is a list and brief definitions of key supply-side programs and demand-side management strategies that HPC data centers can utilize:

– SUPPLY SIDE PROGRAMS

- Peak Load Response: Programs designed for responses during peak hour events and focus on reducing peaks on forecasted high-system load days.
- Reliability Response: These programs are designed for fastest, shortest duration responses. Response is only required during power system availability events that typically cannot be forecasted.
- Regulation Response (Up or Down): Programs designed to follow minute-to-minute commands from the grid to balance the aggregate system load and generation. These programs can be expanded to manage the variable and uncertain generation nature of many renewable resources, supplement the methods for grid scale storage, which is used to store electricity on a large scale to meet the system ramping (up or down) requirements. Pumped-storage hydroelectricity is currently one of the largest forms of grid energy storage. Such responses can also be used for frequency response, which are methods used to keep grid frequency constant and in-balance.
- Congestion Response: Methods used to resolve congestion that occurs when there is not enough transmission capability to support all requests for transmission services. Or, methods used to resolve congestion that occurs when the distribution control system is overloaded. Peak response can be expanded for congestion management.
- Dynamic Pricing: Day-ahead and day-off pricing programs that are designed to change dynamically in response to system load and generation conditions.

– DEMAND SIDE MANAGEMENT STRATEGIES

- Load Shed: Strategies used to reduce the load consumption during peak times, where the reduced load is not used at a later time.
- Load Shift: Strategies where the load during peak times is moved to, typically, non-peak hours.
- Price Response: Strategies for time varying and real-time pricing programs used to motivate modification to electricity consumption.

04

1/13/14
by: sridutt.

05

1/13/14
by: sridutt.

6-7

2 notes:

08

1/13/14

by: sridutt.

- Continuous Energy Management: Emerging strategies that modify the load consumption (increase or decrease) based on intra-hour changing supply conditions.

As the electric grid evolves to a more dynamic and distributed system, its integration with and responses from HPC centers can be used to support supply-side programs.

2.2 Supercomputer Center Response Strategies

One dimension of the model to response to participate in supply-side programs is a list of demand-side management strategies that a supercomputer site might use to manage power in response to a request from their electric service provider.

Although these strategies can be used for managing power in response to a request from an electric service provider, many of them could also be used for improving energy efficiency. It is the former that is of primary interest to this investigation. Two examples may help to clarify this distinction. Load migration is an example of a strategy that is well suited to responding to an electric service provider request, but is not likely to improve energy efficiency. Fine grained power management, on the other hand, is more likely to be used for improving energy efficiency than for responding to electric service provider requests.

Below is a list of strategies:

- Fine grained power management refers to the ability to control HPC system power and energy with tools that are high resolution control and can target specific low level sub-systems. A typical example is voltage and frequency scaling of the CPU.
- Coarse grained power management also refers to the ability to control HPC system power and energy, but contrasts with fine grained power management in that the resolution is low and it is generally done at a more aggregated level. A typical example is power capping.
- Load migration refers to temporarily shifting computing loads from an HPC system in one site to a system in another location.
- Job scheduling refers to the ability to control HPC system power by understanding the power profile of applications and queuing the applications based on those profiles.
- Back-up scheduling refers to deferring data storage processes to off-peak periods.
- Shutdown refers to a graceful shutdown of idle HPC equipment. It usually applies when there is redundancy.
- Lighting control allows for data center lights to be shutdown completely.
- Thermal management is widening temperature set-point ranges and humidity levels for short periods.

09

1/13/14

by: sridutt.

10

1/13/14

by: sridutt.

11

1/13/14

by: sridutt.

12-13

2 notes:

14

1/13/14

by: sridutt.

15

1/13/14

by: sridutt.

3 Prior Work

The prior work described in this paper is that which addresses strategies that HPC centers can take to manage power. A lot of work has been done on energy efficiency, some of which has an element of power management. But, there is not a lot of work that is specifically focused on power management in response to a request from an electrical service provider.

3.1 Fine Grained Power Management

In [12] (Varma et al., 2003) system level DVFS techniques demonstrated. They monitor CPU utilization at regular intervals and then perform dynamic scaling based on their estimate of utilization for the next interval.

Authors in [13] (von Laszewski et al., 2009) present an efficient scheduling algorithm to allocate virtual machines in a DVFS-enabled cluster by dynamically scaling the supplied voltages.

vGreen [3] (Dhiman et al., 2009) is a system for energy efficient computing in virtualized environments by leveraging online workload characterization to dynamic VM scheduling decisions to achieve better performance, energy efficiency and power balance in the system.

3.2 Power Capping

Power capping techniques set a value below the actual peak power and preventing that number from being exceeded through some type of control loop [Fan07]. There are numerous ways to implement this, but they generally consist of a **power monitoring system** such as a power estimation method or one based on direct power sensing, and a **power throttling mechanism**. Power throttling generally works best when there is a set of jobs with loose service level guarantees or low priority that can be forced to reduce consumption when the data center is approaching the power cap value. Power consumption can be reduced simply by **de-scheduling tasks** or by using any available component-level power management **knobs**, such as DVFS [Fan07].

3.3 Job Scheduling

The problem of scheduling jobs has been extensively studied. In general, most of the schedulers implement the First Come First Served (FCFS) policy as a simple but fair strategy for scheduling jobs. But this policy suffers from low system utilization. The most commonly used optimization is backfill [9] [10] [4] [Lif95, Mua95, Fei04]. Backfilling is proposed to improve the system utilization. Backfilling by identifying free capacities allows the smaller jobs fit those capacities to move forward and run on idle processors.

In [Yang ref is not in Zotero] [14] [Yang13] and [Zhou13], job scheduling as a DR strategy and dynamic pricing as a grid integration program have been used

to propose a **power-aware job scheduling** approach to reduce **electricity costs without degrading the system utilization**. The novelty of the proposed job scheduling mechanism is its ability to take *the variation of electricity price* into consideration as a means to make better decisions of the timing of scheduling jobs with diverse power profiles. Experimentations on an IBM Blue Gene/P and a cluster system as well as a case study on Argonne's 48-rack IBM Blue Gene/Q system have demonstrated the effectiveness of this scheduling approach. Preliminary results show a **23%** reduction in electricity cost of HPC systems.

A grid computing infrastructure with large amount of computations normally contains parallel machines (supercomputers cluster) as main computational resources [5] [Fos01]. Incoming jobs to Grid's local resources are scheduled by local scheduling system. **Local scheduling system for parallel machines typically use batch queued space-sharing and its variants as scheduling policies**. Most current local schedulers use backfilling strategies with FCFS queue-priority order as policy for parallel job scheduling.

There are many use cases in grid computing environment that require QoS guarantees in terms of guaranteed response time, including time-critical tasks that must meet a deadline, which would be impossible without a start time guarantee. Furthermore providing time guarantee enable the job to be coordinated with other activities, essential for co-allocation and workflows applications. Advance reservation is a guarantee for the availability of a certain amount of resources to users and applications at specific times in the future [6] [Fos99].

Advance reservation feature requires local scheduling systems to support a reservation capability beside batch queued policy for local and normal jobs. In load migration, we encounter the need to deliver resources at specific times in order to accept jobs from other HPC centers to respond to their demand enforced by electricity grid. This requirement can be achieved by advance reservations [6] [Fos99]. Modern resource management and scheduling systems such as Sun Grid Engine, PBS, OpenPBS, Torque, Maui, and Moab support backfilling and advance reservation capabilities.

3.4 Load Migration

In order to balance the electrical grid, [2] [Chiu12] proposes a low-cost **geographic load migration** to match electricity supply. In addition, authors present a real grid balancing problem experienced in the Pacific Northwest. They propose a symbiotic relationship between data centers and electrical grid operators by showing that **mutual cost benefits** can be accessible.

3.5 Thermal Management

Thermal and cooling metrics are becoming important metrics in scheduling and resource management of HPC centers. **Runtime cooling strategies are mostly job-placement-centric**. These techniques either aim to place incoming computationally intensive jobs in a thermal-aware manner on servers with lower tempera-

19

1/13/14
by: sridutt.

20

1/13/14
by: sridutt.

21

1/13/14
by: sridutt.

22

1/13/14

by: sridutt.

tures or attempt to reactively migrate/load-balance jobs from high temperature servers to servers with lower temperatures. T^*

[8] [Kau12] takes a data-centric thermal- and energy-management approach and does proactive, thermal-aware file placement which allows cooling energy costs savings without performance trade-offs. T^* is cognizant of the uneven thermal-profile of the servers, differences in their thermal-reliability-driven load thresholds, and differences in the data-semantics, i.e., computation job rates, sizes, and evolution life spans, of the big data placed in the cluster.

In this paper, we assume that the grid is a given constant as a fundamental property. But, grid integration solutions may take into consideration that it isn't a given as electrical grid infrastructures will evolve in the future [7] [He08]. Thus, changes in the grid could make grid integration more or less difficult.

In [1] [Aik11], authors explored the potential for HPC centers to adapt to dynamic electrical prices, variation in carbon intensity within an electrical grid, or availability of local renewables. Through simulations experiments on workloads from the Parallel Workloads Archive alongside real-world pricing data, they demonstrate potential savings on the cost of electricity ranging typically between 10-50%. Nonetheless, adaptation to the variation in the electrical grid carbon intensity was not as successful, but adaptation to the availability of local renewables showed potential to significantly increase their use.

23-24

2 notes:

25

1/13/14

by: sridutt.

4 Questionnaire

We used a questionnaire in order to understand the current experiences of supercomputer center's interaction with their electricity service providers. We restricted the analysis to sites in the United States since the results of the survey and practices of demand response is highly correlated and driven by energy policies in the country. [11] [Tor10].

Nineteen Top100 List sized sites in the United States were targeted for the questionnaire. Eleven sites responded (Oak Ridge National Laboratory, Lawrence Livermore National Laboratory, Argonne National Laboratory, Los Alamos National Laboratory, LBNL, Wright Patterson Air Force Base, National Oceanic Atmospheric Administration, National Center for Supercomputing Applications, San Diego Supercomputing Center, Purdue University and Intel Corporation). Eight sites didn't respond (National Center for Atmospheric Research, IBM Corporation, National Energy Technology Laboratory, Indiana University, Texas Advanced Computing Center, Sandia National Laboratory, National Renewable Energy Laboratory, National Aeronautics and Space Administration). The questionnaire was sent to a sample that was not randomly selected. It was sent to those sites where it was relatively easy to identify an individual based on membership within the EE HPC WG. The sample is more representative of Top50 sized sites (1 Top50 sized site was not in the sample and 60% (9/15) of the sample responded). Only 4 additional sites were sampled from the Top51-Top100 List and, of those, 2 responded (Intel and National Oceanic and Atmospheric Administration).

The total power load as well as the intra-hour fluctuation of these sites varied significantly. There were four sites with total power load greater than 10MW, two sites with ~5MW total power load and five sites with less than 2MW total power load. We chose less than 3MW intra hour variability as the bottom of the scale because we assumed that the electrical service providers would not be affected by 3MW (or less) fluctuations. For those with total power load greater than 10MW, the intra-hour fluctuation varied from less than 3MW to 8MW. One of ~5MW sites said that they experienced 4MW variability. The rest of the sites were all less than 3MW. Most of the intra-hour variability was due to preventative maintenance.

26
1/13/14
by: sridutt.

Table 1. Caption Number 1

Total Load	Variability	Frequency
16-17MW	5MW	weekly
13-14MW	8MW	monthly
10-11MW	Less than 3MW	weekly
10-11MW	7MW	weekly
4-5MW	Less than 3MW	weekly
4-5MW	4MW	weekly
1-2MW	Less than 3MW	weekly
1-2MW	140kW	daily
1-2MW	Less than 3MW	yearly
1-2MW	200kW or less	daily
1-2MW	Less than 3MW	daily

We asked if the supercomputer centers had talked to their electric service providers about programs and methods used to balance the grid supply and demand of electricity. About half of them have had some discussion, but it has mostly been limited to demand side and not supply-side programs.

Table 2. Caption Number 2

Discussions with Electricity Providers	% Answered Yes
Demand-side programs	
Shedding load during peak demand	54
Responding to pricing incentive programs	45
Shifting load during peak demand	36
Supply-side programs	
Enabling use of renewables	36
Congestion, Regulation, Frequency Response	18
Contributing to electrical grid storage	10

27

1/13/14
by: sridutt.

More than half of the respondents are not interested in shedding or shifting load during peak demand. There is some indication that this low interest is primarily due to the lack of a clear business case for the sites where there is interest, shifting is more attractive than shedding load. The San Diego Supercomputing Center (SDSC) is an exception to this trend because of a site-wide program. "University of California San Diego generates 30-35MW of power yet still imports 5-10MW. As a large generation source the utility providers see the campus as a highly attractive partner for offloading grid stress. Automatic load shedding is being explored/deployed today."

Responding to pricing incentive programs is also not considered interesting, although the reasons for this low interest may be organizational. Several open-ended comments revealed that pricing is fixed and/or done by another organization at the site level and outside of their immediate control.

Eighty percent of the respondents have not had discussions with their electricity service providers about congestion, regulation and frequency response. Los Alamos National Laboratory (LANL) is one of two who have had discussions and who commented that they are "learning about the process" and that it is "outside of [their]

visibility or control".

There were been many more respondents who have had discussions with their electricity service providers about enabling the use of renewables; 36% have already had discussions and more than half are interested in further and/or future discussions. SDSC already has a site-wide program; "the campus has a large fuel cell (2.5+ MW) and works with the utility with renewables." Other responses suggest that the interest is at the site level and not unique to the supercomputer center.

An open-ended question was posed as to whether or not there was information either requested of the supercomputer sites by their providers or, conversely, requested of the providers by the sites. In both cases, well over 75% of the respondents answered no. Lawrence Livermore National Laboratory (LLNL) and LANL were the exceptions. LLNL is "working on obtaining additional data from them and a means of sharing data between them and us" and has been requested to provide "additional detailed forecasting and ultimately real time data." LANL has also been requested to provide "power projections, hour by hour, for at least a day in advance" and, perhaps as a consequence, would like to have more information on "sensitivity of power distribution grid to rapid transients (random daily step changes of 10 MW up or down within a single A cycle)."

Given the low levels of current engagement between the electricity service providers and the supercomputer centers, it is not surprising that none of the supercomputer centers are currently using any power management strategies to respond to grid requests by their electrical service providers. SDSC's *supercomputer center* is not an exception, but they did respond that their entire "campus is leveraging parallel electrical distribution to trigger diesel generators and other back-up resources to respond to to grid and non-grid requests."

28

1/13/14
by: sridutt.

We tried to evaluate if power management strategies will be considered relevant and effective for grid integration at some point in the future. Two questions were asked; is there interest in using the strategies and what impact did they think that the strategies would have. When combining interest and impact, the results showed that power capping, shutdown, and job scheduling were both high interest and impact. Load migration, back-up scheduling, fine grained power management and thermal management were medium interest and impact. Lighting control and back-up resources were low interest and impact.

Table 3. Caption Number 3

HPC strategies for responding to Electricity Provider requests (listed from highest to lowest interest + impact)
Coarse grained power management
Facility shutdown
Job scheduling
Load migration
Re-scheduling back-ups
Fine grained power management
Temperature control beyond ASHRAE limits
Turn off lighting
Use back-up resources (e.g., generators)

5 Opportunities/Solutions and Barriers

The biggest opportunity in the integration of the electrical grid and supercomputer centers is to start a process of negotiation/interaction between utility providers and HPC centers; the survey data indicates that this is being asked for by providers, at least in some small way already.

SDSC stands out as a leader in integrating with their utility provider. They have implemented a site-wide integration. A probably opportunity for other supercomputer centers would be to pursue a similar site-wide integration.

The two largest power draw sites (LLNL and ORNL) seem to have divergent experiences. Not clear why, but probably has more to do with their providers than their sites. LLNLs experience may be seen with other sites. LANL seems to be moving in that direction. This may portend a required opportunity for some sites.

There isn't a clear business case on the part of the supercomputing center for pursuing integration. This probably needs to be further understood and developed.

According to the survey data and literature review, the greatest opportunities for HPC centers to develop integration capabilities are related to system software. System software in these areas in the order of their interest and impact

consist of coarse grained power management (power capping), job scheduling, load migration, rescheduling back-ups and fine grained power management.

Sites are developing experience with energy efficiency that can transfer to power management for utility integration.

Job scheduling is seen as having the greatest interest and impact- because from a practical perspective, coarse grained power management is dependent upon job scheduling.

If integration starts happening between utility providers and HPC centers, the system software (i.e., job scheduler) is a key component in order to ensure that this happens as efficiently as possible in order to keep high utilization / business utility going at the HPC center. (Consider here things such as fluctuations in HPC use; e.g., things like large-scale acceptance / Top500 style runs). Load migration requires cooperation between HPC centers (and also may require overcoming difficult challenges like migrating large datasets, ensuring security requirements are met and porting tuned code). By using *advance reservation capabilities* of schedulers (within local resource managers) of HPC centers, we facilitate the execution of load migration strategy between HPC centers (e.g., in terms of automation); as a result we increase the interest level and to some extent the impact level of load migration strategies.

In addition to system software, infrastructure software such as grid computing may have a place in opportunities. In the US, supercomputer centers are connected via grid computing infrastructures such as TeraGrid, Open Science Grid. Grid computing's protocols, interfaces, and standards can facilitate the execution of DR strategies, as a result grid computing may increase the interest level and/or the impact level of DR strategies.

If there is some kind of automated "negotiation" process that takes place between utility providers and HPC centers, it's likely that the utility providers will need to improve their capabilities to be able to participate in this process (e.g., probably need to solve some kind of weighted optimization problem in near real time in order to know where the most important places are to ensure uninterrupted service); there is an opportunity to the utility provider in this, however, in that these advancements in their technology for monitoring and adjusting their infrastructure might be leveraged toward other ends that are not related to HPC centers specifically.

The survey data indicates that grid programs are being negotiated between utility providers and HPC centers more than that of grid methods. Due to the lack of a clear business case there is low interest in shedding and shifting load during peak demands. Nonetheless, according to the survey data in HPC landscape load shifting is more attractive than shedding load. In addition, there are high interests and opportunities in the use of renewables at the site level, according to the survey data and literature review. According to literature review, there are great opportunities in terms of reducing electricity costs in exploiting dynamic pricing as a grid integration program.

In contrary, there is lack of knowledge on the use and integration aspects of congestion, regulation, and frequency response methods.

31

1/13/14
by: sridutt.

32

1/13/14
by: sridutt.

There are needs and opportunities for the supercomputer sites in having accounting system. According to the survey data, there are information request on electricity usage by electricity providers from the supercomputer sites for getting detailed forecasting and real time data. Accounting data can be used to forecast and model future energy usage of an HPC center. So this can be communicated and be integrated with electricity grid. Accounting data can be classified in terms of HPC center components, cooling, systems, lighting, etc. If/how electricity grid providers can use energy and usage accounting data to plan electricity provisioning of an HPC center?

The HPC/electrical grid integration at the communication and message level need to be standardized and be interoperable. Interfaces, communication infrastructure, data, information exchange, and agreement should be based on standards.

6 Conclusions and Next Steps

1. Potential HPC-specific value proposition for active DR engagement
2. Based on Grid Integration solutions – local and system-wide impacts
3. Next steps – specific directions or target areas to focus

Electricity providers have viewed the hourly, daily, and seasonal fluctuations of demand as facts of life. These fluctuations required additional generating capacity, particularly peaking plants that were needed only a few hours per year.

DR adoption: automation

automation technologies

DR market:

value proposition,

the measurement and verification models,

patents,

intellectual property

lighting, and heating, ventilation, and air conditioning (HVAC)

electricity-price markets

interoperability

The grid integration need to be standardized and provide interoperable interfaces to be interoperable. Interfaces, communication infrastructure, data, information exchange, agreement should be based on standards. Communication with grid providers need to be standardized. grid request/response messages. Requests include DR event, price, renewable generation

How is architected an accounting system (energy and utilization) of an HPC center? based on sensor systems like in [Hay09] . Sensor systems for an HPC center to report real time power consumption of various components such as cooling, compute systems, storage, networks, racks, etc.

[Hay09] S. Hay and A. Rice, “The case for apportionment,” in Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, New York, NY, USA, 2009, pp. 13–18.

Apportioning the total energy consumption of a building or organisation to individual users may provide incentives to make reductions. We explore how sensor systems installed in many buildings today can be used to apportion energy consumption between users. We investigate the differences between a number of possible policies to evaluate the case for apportionment based on energy and usage data collected over the course of a year. We also study the additional possibilities offered by more fine-grained data with reference to case studies for specific shared resources, and discuss the potential and challenges for future sensor systems in this area.

- If accounting data can be used to forecast and model future energy usage of an HPC center? so this can be communicated and be integrated with electricity grid.
- If/how electricity grid providers can use energy and usage accounting data to plan electricity provisioning of an HPC center?
- user-specific accounting data versus workload-specific accounting data.
- accounting data in terms of HPC center components, cooling, systems, lighting, etc.

These are excellent questions. What you've outlined below is a set of value of real-time data (the term "accounting" confused me earlier) of energy and utilization for HPC systems. Some of these values are for EE and the rest is how the electric grid service providers can benefit from it. For example, telemetry data for wholesale DR markets and M&V.

M&V or Measurement and Verification refers to quantification of load shed that a particular load is participating in. Typically, there are many baseline methodologies that the utilities and ISOs use to calculate the amount of DR a particular load/facility is providing through real-time and day-ahead metered data. The metering and telemetry to provide the M&V is key in determining if a particular resource can participate in a DR market and validate its performance for settlement (economics).

References

1. D. Aikema and R. Simmonds. Electrical cost savings and clean energy usage potential for HPC workloads. In *2011 IEEE International Symposium on Sustainable Systems and Technology (ISSST)*, pages 1–6, 2011.
2. D. Chiu, C. Stewart, and B. McManus. Electric grid balancing through lowcost workload migration. *SIGMETRICS Perform. Eval. Rev.*, 40(3):4852, Jan. 2012.
3. G. Dhiman, G. Marchetti, and T. Rosing. vGreen: a system for energy efficient computing in virtualized environments. In *Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design, ISLPED '09*, page 243248, New York, NY, USA, 2009. ACM.
4. D. G. Feitelson, U. Schwiegelshohn, and L. Rudolph. Parallel job scheduling - a status report. In *In Lecture Notes in Computer Science*, page 116. Springer-Verlag, 2004.
5. I. Foster. The anatomy of the grid: enabling scalable virtual organizations. In *First IEEE/ACM International Symposium on Cluster Computing and the Grid, 2001. Proceedings*, pages 6–7, 2001.
6. I. Foster, C. Kesselman, C. Lee, B. Lindell, K. Nahrstedt, and A. Roy. A distributed resource management architecture that supports advance reservations and co-allocation. In *1999 Seventh International Workshop on Quality of Service, 1999. IWQoS '99*, pages 27–36, 1999.
7. M. M. He, E. M. Reutzel, X. Jiang, Y. H. Katz, S. R. S, and D. E. Culler. An architecture for local energy generation, distribution, and sharing. In *IEEE Energy2030 Conference Proceedings*, Atlanta, Georgia, USA, Nov. 2008.
8. R. T. Kaushik and K. Nahrstedt. T*: a data-centric cooling energy costs reduction approach for big data analytics cloud. SC '12, page 52:152:11, Los Alamitos, CA, USA, 2012. IEEE Computer Society Press.
9. D. A. Lifka. The ANL/IBM SP scheduling system. In *In Job Scheduling Strategies for Parallel Processing*, page 295303. Springer-Verlag, 1995.
10. A. W. Mu'alem and D. G. Feitelson. Utilization, predictability, workloads, and user runtime estimates in scheduling the IBM SP2 with backfilling. *IEEE Trans. Parallel Distrib. Syst.*, 12(6):529543, June 2001.
11. J. Torriti, M. G. Hassan, and M. Leach. Demand response experience in europe: Policies, programmes and implementation. *Energy*, 35(4):1575–1583, Apr. 2010.
12. A. Varma, B. Ganesh, M. Sen, S. R. Choudhury, L. Srinivasan, and B. Jacob. A control-theoretic approach to dynamic voltage scheduling. In *Proceedings of the 2003 international conference on Compilers, architecture and synthesis for embedded systems, CASES '03*, page 255266, New York, NY, USA, 2003. ACM.
13. G. von Laszewski, L. Wang, A. Younge, and X. He. Power-aware scheduling of virtual machines in DVFS-enabled clusters. In *IEEE International Conference on Cluster Computing and Workshops, 2009. CLUSTER '09*, pages 1–10, 2009.
14. Z. Zhou, Z. Lan, W. Tang, and N. Desai. Reducing energy costs for IBM blue Gene/P via power-aware job scheduling.

7 Appendices

7.1 Background

Over the last few years, load growth, increases in intermittent generation, declining technology costs and increasing recognition of the importance of customer

behaviour in energy markets have brought about a change in the focus of DR in Europe [Tor10]. The long standing programmes involving large industries, through interruptible tariffs and time of day pricing, have been increasingly complemented by programmes aimed at commercial and residential customer groups.

[11] Tor10] examines experiences within European countries as well as at European Union (EU) level. While business programmes, technical and economic potentials vary across Europe, there are common reasons as to why coordinated DR policies have been slow to emerge: the limited knowledge on DR energy saving capacities; high cost estimates for DR technologies and infrastructures; and policies focused on creating the conditions for liberalising the EU energy markets.

advances in DR It describes initiatives, studies and policies of various European countries, with in-depth case studies of the UK, Italy and Spain.

Spees, K., & Lave, L. B. (2007). Demand Response and Electricity Market Efficiency. *The Electricity Journal*, 20(3), 69–85. doi:10.1016/j.tej.2007.01.006

Interruptible Programmes represent 6.5% of peak power and Load Shedding Programmes initiate automatic load shedding in emergency situations [30]. With Interruptible Programmes participants are required to reduce their load to pre-defined values. With Load Shedding Programmes utilities have the possibility to remotely shutdown participants' equipment at short notice. One significant difference between these two programmes is that for Interruptible Programmes participants who do not respond can face penalties.

Load Shedding Programmes are divided into real time programmes (without notice) and 15 min notice programmes. The size ranges from 1200 MW for real time programmes to 1750 MW for notice programmes. Participants in these programmes have to install and maintain Load Shedding Peripheral Units and will be compensated according to a non-market price defined in regulation. The size of curtailable power is of 10 MW for programmes without notice and 3 MW for programmes with notice.

Load forecasting is very important for power system operation and planning. Traditional load forecasting tools have limitations to reflect DR customer behaviors into load predictions. In [Zhou12], existing DR contracts are reviewed for both wholesale and retail markets. In this study, an illustrative example is provided to explore the impact of these contracts on load forecasting. In conclusion, a concept of proactive load forecasting based on contract types is proposed for forecasting loads in a smart grid environment.

Modern hardware components such as processor, memory, disk and network offer feature sets (Burd and Brodersen, 1995) to support energy aware operations. Exploiting these feature sets in order to be more energy efficient is a very important and challenging task in modeling cost/performance trade-offs, in designing algorithms, and in defining policies. Today processors offer two power-aware features, i.e., cpuidle and Dynamic Voltage and Frequency Scaling (DVFS). The cpuidle feature offers a number of CPU power states (C-states) in which they could reduce power when CPU is idle by closing some internal gates.

The CPU C-states are C0, C1, ..., Cn. C0 is the normal working state where CPU will execute instruction, and C1, ..., Cn are sleeping state where CPU stops executing instruction and power down some internal components to save power. The DVFS is another power-saving method especially when CPUs are in load line, allowing quick adjustment to frequency/voltage upon demand in small interval. The key idea behind DVFS techniques is to dynamically scale the supply voltage level of the CPU so as to provide just-enough circuit speed to process the system workload, thereby reducing the energy consumption.

We received eleven responses from the following sites: LLNL, ANL, Intel, SDSC, Illinois, NOAA, ORNL, Purdue, AF, NERSC, LANL. The table below shows the ranks of the responding organization in the top 100 list. 80% of the top 10, 60% of the top 50, and 30% of the top 100 participated in the survey.

Table 4. Participating organizations and their rank in the top 100.

Organization	ORNL	LLNL	ANL	LANL	NERSC	Purdue	AF	NOAA	Intel	SDSC	Illinois
Rank	2	3	5	22	24	28	40	48	71	102	

The survey had a total of 29 questions, eight of the questions were skipped 16 times by the 11 participants. This is only 5% of the total questions answered by the eleven participants. The graph below shows the distribution of skipped questions.

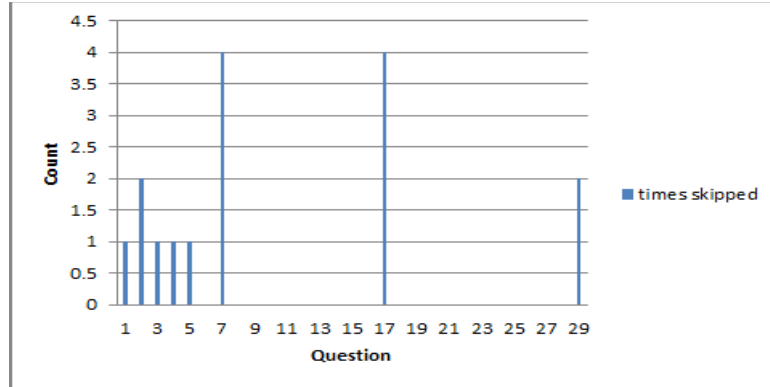


Fig. 1. Figure 2

Question 7 and 17 were skipped 4 times each, these two questions are follow up questions to questions 6 and 16 and we were trying to collect more details about user's responses in questions 6 and 16.

The Electrical Grid and Supercomputer Centers : An Investigative Analysis of Emerging Opportunities and Challenges

Lname, Fname

- | | | |
|--|-----------------------------|------------|
| 01 | sridutt bhalachandra | Page no. 1 |
| 11/1/2014 1:27 | | |
| Rephrase? | | |
| 02 | sridutt bhalachandra | Page no. 1 |
| 11/1/2014 1:29 | | |
| Do we have the study that contains these values? | | |
| 03 | sridutt bhalachandra | Page no. 1 |
| 11/1/2014 1:30 | | |
| Cite | | |
| 04 | sridutt bhalachandra | Page no. 3 |
| 13/1/2014 15:31 | | |
| How are energy conservation and energy efficiency supply programs different? | | |
| 05 | sridutt bhalachandra | Page no. 3 |
| 13/1/2014 15:36 | | |
| Can be rephrased? | | |
| 06 | sridutt bhalachandra | Page no. 3 |
| 13/1/2014 15:42 | | |
| The description of both demand and supply programs need to better structured using bulleted lists. | | |
| 07 | sridutt bhalachandra | Page no. 3 |
| 13/1/2014 15:44 | | |
| Can be rephrased? Would it mean not allow any more jobs to be scheduled? | | |
| 08 | sridutt bhalachandra | Page no. 4 |
| 13/1/2014 15:53 | | |
| Arent Price response and Continuous energy mgmt essentially same? | | |

13/1/2014 16:01

Rephrase entire sentence, cannot clearly understand. Could be "One dimension of the response model, to participate in supply-side programs is through a list of demand-side management strategies. These strategies can be used by a supercomputer site to manage power in response to a request from their electric service provider."

13/1/2014 16:13

Clarification is understandable but why are we clarifying it? Whats the motive to put this clarification. Am I missing something?

13/1/2014 16:15

Again can be better organized with bulleted list.

13/1/2014 16:48

Is this against the premise for HPC applications that we make in the introduction saying the applications are not portable due to security, data-locality and system tuning?

13/1/2014 16:51

May not be a relevant question now but for future, how to we understand the power profile of the applications?

13/1/2014 16:55

Shutdown depends on the utilization of facility. Thus dynamically shutting a fraction of the facility may not uphold the stable power demand that we want from the electric grid. I understand this comment is too specific but I am not sure to what level of detail we would be going in this paper.

13/1/2014 16:58

I am still not sure if the organization of this section is right. We begin the section saying that we would be using an integration model that has been used for datacenters and would be tweaking it for HPC. I dont feel that we have developed on this premise and the sections seems to be too loosely written.

13/1/2014 17:18

lot of work?

13/1/2014 17:36

The two references above are excellent works but take into account virtualized environments. But it is said in the intro "As opposed to data centers, supercomputer centers have very high system utilization and are not likely to use virtualization as a strategy." This I understand can be extrapolated to SC facility. But there are more relevant papers like Just in Time DVFS and others that can be cited. Also other works on fine grain power mgmt like duty cycle modulation, clock gating etc. also exists.

13/1/2014 17:43

1. The last line seems inappropriate here? Any scheduling term should be in the next subsection? 2. Power capping tries to solve the problems of DVFS which requires accurate estimation of optimal frequency. If we say we will use DVFS again for power mgmt it wouldnt be any good. (I am aware power capping uses DVFS internally, but I think this can be rephrased.) Citing Barry's work Beyond DVFS too would be a good option.

13/1/2014 18:03

Add in next subsection?

13/1/2014 18:12

This subsection needs to be better organized. 1. Paragraph 3 needs to be in para1 where we say what the present job scheduling strategies are. 2. Then we use the first paragraph to talk about studies on FCFS and backfill. 3. Then we use the last para speaking about the QoS and advance reservations. Also, Slurm scheduler is quite popular now and it can be added to the list of schedulers 4. We speak Yang13 and how it uses the electricity costs. 5. There is another paper by Yang that appeared in SC13 titled "Integrating Dynamic Pricing of Electricity into Energy Aware Scheduling for HPC Systems". I am not sure if both papers are same but that too had some good results. I had just attended talk but I havent read the paper.

13/1/2014 18:17

1. Is using word "datacenter" correct? I understand HPC applications tend to be I/O intensive too. And recently big-data and hpc have converged a lot. But I wasn't sure. 2. Again, is load migration right for HPC? (wrt premise used in the Introduction)

13/1/2014 18:19

This migration I understand is within the SC facility. But is it a viable option with concern to data locality and other issues?

23	sridutt bhalachandra	Page no. 7
	13/1/2014 18:29	
	What will be the constants?	
24	sridutt bhalachandra	Page no. 7
	13/1/2014 18:30	
	Cant clearly understand. Rephrase may be?	
25	sridutt bhalachandra	Page no. 7
	13/1/2014 18:35	
	How is this relevant to thermal mgmt and in this subsection? Is it by relation to carbon intensity?	
26	sridutt bhalachandra	Page no. 8
	13/1/2014 19:34	
	Is there a basis of this assumption if we are questioned? - intra-hour?	
27	sridutt bhalachandra	Page no. 9
	13/1/2014 19:52	
	Can we add what is this indication? Is this remark based on the SDSC case that follows?	
28	sridutt bhalachandra	Page no. 9
	13/1/2014 20:03	
	When we get this data, it will be really interesting	
29	sridutt bhalachandra	Page no. 10
	13/1/2014 20:48	
	Rephrase sentence unclear.	
30	sridutt bhalachandra	Page no. 10
	13/1/2014 20:50	
	portend is a negative word and signifies unpleasant. harbinger a better word? Frankly I am not getting the meaning of the sentence, so portend may be right too.	
31	sridutt bhalachandra	Page no. 11
	13/1/2014 20:58	
	More elaboration?	

13/1/2014 21:06

Rephrase, not clearly understandable.