The Science of Measurement:

Improving Data Center Performance with Continuous Monitoring and Measurement of Site Infrastructure

John Stanley
Jonathan Koomey
October 28, 2009

Table of Contents

Executive Summary

This report explains some of the key considerations surrounding performance monitoring and measurements in data centers. The key lessons for data center managers are:

- Measurement is conducted to inform decisions on actions that can enhance a data center's business value by better accommodating growth, reducing costs, or increasing uptime.
- There are opportunities to improve efficiency through a continuous process of incremental changes, as well as through the more traditional "one shot" project-based approach.
- The business or strategic objectives of the measurement effort should drive measurement decisions, and having "the right data" is more important than simply having "more data"
- Data center managers should think carefully before beginning a monitoring program, so that the data collected will be maximally useful for judging operational performance and making decisions.
- After collecting data, managers should take action based on what they have learned. Managers should then continue collecting data to get feedback on the effectiveness of their actions.
- When evaluating potential monitoring systems, managers should examine multiple factors, including:

    - ◊ Ability to collect data from all desired devices
    - ◊ Granularity of data collection
    - ◊ User friendliness and ease of integrating data across devices and time scales
    - ◊ Scalability for mass deployment and multi-site capability
    - ◊ Adaptability to new measurement needs
    - ◊ Trending and analysis of data
    - ◊ Integration with control systems
    - ◊ Ability to detect problems and notify data center operators

Goal and structure of this report

The goal of this report is to help data center managers (both IT and facilities managers) better understand the business value of measuring data center performance characteristics and the considerations that are important to developing a measurement scheme. Many of the concepts and anecdotes in this report draw from three case studies in the *Measurement Series*, published by Modius, Inc., in which commercial data center monitoring software was deployed in an operating data center environment.[1] This report can be read as a stand-alone document, but readers are encouraged to refer to the *Measurement Series* case studies for additional information.

The case studies are:

- *Case Study: Power Chain Capacity Expansion at Sybase*, in which current draw in power distribution units was monitored to enable reliability improvements
- *Case Study: Cooling & Chilled Water Efficiency Project at Sybase*, in which cooling systems were monitored to reduce energy use and improve reliability
- *Case Study: Availability Improvements Using Granular Monitoring at Sybase*, in which monitoring of data center health helped intercept problems before they could cause downtime

Full citations for these case studies can be found in the References section.

Why measure – The business value of measuring data center performance

Different types of business value

In a data center, measurement is not just an academic exercise; managers collect data so they can make decisions and take action. Data gathering that does not support these activities is a waste of time and money. Data center monitoring can be helpful to managers in three strategic areas:

Maximizing data center capacity and accommodating growth

In most companies, the demand for data center services is growing rapidly. These services are often either direct revenue generators or they are critical to business support functions such as product development, resource planning, or communication. No data center manager wants to be caught unable to add Information Technology (IT) equipment because the data center is suddenly out of space, power, or cooling capacity. Monitoring can help managers plot long term growth trends and anticipate future facility needs.

Minimizing Total Cost of Ownership (TCO)

TCO is the total lifetime cost of owning and operating a data center, including the IT equipment, software, maintenance, building lease or depreciation costs, power and cooling equipment, electricity, and all other expenditures.[2] The business goal is to provide the best data center service (# of end-users served, # of computations, etc.) for the lowest cost. Effective monitoring can drive down TCO by improving energy efficiency and by helping capacity planners minimize the costs of unnecessary over-provisioning.

[1] Modius, Inc. is a firm that develops intelligent measurement systems for mission-critical facilities. See <http://www.modius.com/>.

[2] See *A Simple Model for Determining True Total Cost of Ownership for Data Centers* by Koomey et. al. (2007) for an extensively documented breakdown of capital and operating costs for a "typical" data center. In particular, readers will note that site infrastructure capital costs and energy costs are significant parts of TCO.

Maximizing availability and business continuity

> Data center managers also know that uptime is critical. Unplanned (or even planned) outages represent time when revenue-generating services are offline or other critical business functions cannot be performed. Good data center monitoring can improve uptime by allowing managers to detect emerging problems before they cause outages.

Granular monitoring of data center performance variables can allow managers to more effectively diagnose problems and get the most out of their facilities. For example, many facilities have "hot spots" with extra cooling needs, and these needs may even change throughout the course of the day as IT equipment ramps up and down. In the past, data center managers sometimes had to "blindly" over-provision cooling throughout the facility and hope that the problem areas (wherever they were) could be controlled. With detailed monitoring data, managers can now *find* the hot spots, *act* to cool them appropriately, and *verify* that the new cooling was effective. Also, managers do not need to waste money on extra cooling units for other parts of the data center, and they can continue to add productive IT equipment as needed without the fear that some part of the facility might unexpectedly get too hot.

There are hundreds or thousands of potential operational variables to measure in a data center, from power draw to air temperature to water flow. Managers can use these data to inform decisions on actions they might take to enhance their business objectives of raising productivity, reducing costs, and improving uptime.

Capturing value via an incremental approach to improvement

In some data centers, noteworthy improvements in infrastructure efficiency occur only during large, "one shot" projects such as major retrofits or efficiency overhauls. However, there are also significant opportunities to improve data center performance through a continuous, data-driven process of incremental changes.

A recent report by a leading IT research and advisory company recommends that data center operators adopt a continuous-improvement approach to data center optimization rather than a "one shot" project-based approach. The report also notes that this continuous approach fits well with models such as Information Technology Infrastructure Library (ITIL) and Capability Maturity Model Integration (CMMI).[3] [4]

Illustrating the value of such an iterative approach, one data center was able to realize significant energy and financial savings by raising the chilled water temperature of its cooling system by 9 ºF over a period of about two years. The data center operators used monitoring data to locate and correct hot-spots, easing the burden on the cooling system and allowing small, controlled increases in the chilled water temperature. Repeating this iterative process over a period of two years, they were able to raise the chilled water temperature from 43 ºF to 52 ºF.[5]

Note, however, that incremental improvement schemes still require data center managers to use the same diligence as project-based approaches in order to maintain reliability at mission critical sites. Under either strategy, changes still need thorough analysis, meticulous planning, and careful execution.

---

[3] Kumar 2009, p.4.

[4] Information Technology Infrastructure Library (ITIL) is a popular approach and set of best practices for IT service management – see <http://www.itil-officialsite.com>. Capability Maturity Model Integration (CMMI) is another popular approach – see <http://www.sei.cmu.edu/cmmi/>.

[5] See *Case Study: Cooling & Chilled Water Efficiency Project at Sybase*, p.6-8, 17.

How to measure – General lessons on monitoring and data collection

The multi-layered view of measurement

Data center monitoring can be viewed as a multi-layered process. The highest layer contains the business objectives that data center managers are trying to achieve – meeting these objectives is the whole reason for measuring at all. Subsequent layers descend into more technical details, with the lowest layer describing the characteristics of the actual sensors that do the data gathering. **Figure 1** below shows these layers.

Layers of the Measurement Process

INCREASING PROXIMITY TO DATA

Layer	Key question	Example
Strategic / business objectives	*What business outcome do we want to accomplish with our measurement scheme?*	Improve uptime, reduce cost
Critical thinking	*What needs to be measured to contribute to our objectives?*	Weekly actual power consumption vs. available capacity, rack-by-rack hourly temperatures
High-level monitoring system capabilities	*What makes a good monitoring system overall?*	Scalability, user-friendliness, ability to aggregate data from many sources
Low-level monitoring system capabilities	*What technical details matter?*	Granularity of data, error reporting, robustness
Sensors	*What's important about the sensors collecting the raw data?*	Resolution, drift

Figure 1 – *The layers of the measurement process*

Think before you measure

Key points:

- Business / strategic objectives of the measurement effort should drive measurement decisions
- Having "the right data" is more important than simply having "more data"
- Think critically about what data are needed and why
- Know the limits of the data
- Normalize measurements for external factors
- Measure what's important

Business or strategic objectives of the measurement effort should drive measurement decisions

Before beginning any data center monitoring initiative, managers should start by reviewing the business or strategic objectives of the effort. Is the goal of making measurements to assist in capacity planning? To reduce TCO? To provide savings documentation for a utility rebate or public environmental initiative? To improve reliability?

The objectives provide the grounding for the rest of the thinking below. Also, when faced with choices regarding the measurement plan, data center managers can ask themselves "If we did X, how would it contribute to our goals?" This is often a useful test. Finally, note that objectives can change or expand over time, and measurement plans may need to be adapted. However, outlining the goals in as much detail as possible at the beginning is useful.

Having "the right data" is more important than simply having "more data"

"More data" is not always better if it does not tell a manager what he or she wants to know. This is worth remembering as devices, monitoring systems, and sensors continue to add new features and data collection abilities. Obviously, getting the data you *do* need is critical. Thinking through what data is "the right data" involves several steps, as described below.

Think critically about what data are needed and why

Data center managers who start measuring before they start thinking risk spending a lot of money on sensors and systems collecting data they will never use. Even worse, they might realize after a few months of data collection that the information they have is not what they need to reach their goal. For example, some large data centers perform their own measurements of power draw (rather than trusting manufacturer data) on new server models in order to determine how much Uninterruptible Power Supply (UPS) capacity[6] they will need in order to support hundreds or thousands of such servers. Imagine a manager running a week of tests before realizing that the power meters were collecting data on *average* power draw instead of (the higher and more relevant) *peak* power draw—in this case, the data collected might be virtually useless.

Also, data center managers should consider not only what data *they* would like to see, but also what data would be useful to convince other decision makers to authorize necessary action. For example, at one data center firm, solid trend data on steadily growing power use played a key role in convincing upper management that undertaking a large facility upgrade would be worthwhile.[7]

[6] A UPS system conditions raw utility power coming into a data center by filtering out voltage sags, spikes, and other noise. UPS systems also contain battery (or other) backup to keep continuous power flowing to IT equipment during short utility outages, and they can seamlessly switch over to draw power from generators during longer utility outages. The capacity of a UPS system is typically rated in the number of kVA or kW it can provide to the IT equipment connected to it.

[7] See *Case Study: Power Chain Capacity Expansion at Sybase*, p.12.

Having a simple model of cause and effect in the data center can be helpful in deciding what needs to be measured. Knowing that increases in IT load naturally drive increases in cooling demand is important, if one hopes to measure cooling efficiency improvements. For reliability-focused measurement initiatives, an operator's knowledge of data center problems and their symptoms can provide direction on what data to collect.

Thinking about what data is necessary will also help managers make informed decisions about the required precision of measurements (*e.g.,* +/- 1 ºF or +/- 0.01 ºF) and about the appropriate level of data aggregation (each server individually, the average of all servers, etc.).

Finally, managers should also consider collecting data that they think *might* be needed in the future, even if it is not immediately necessary. Later reliability or efficiency initiatives could require historical data for trend analysis or determining "baseline" performance.

Know the limits of what can be concluded from the data to be gathered

Data center managers should also think about the limits of what they will be able to reasonably conclude from the data they collect. As mentioned in the example above on server power measurement, data on average power draw are not useful for making conclusions about how much UPS capacity will be necessary for mass deployment of a particular server model. Peak power measurements are more important. (However, data on average power draw *are* still useful for estimating things like annual energy costs.)

Similarly, what can a data center manager conclude by going to two facilities and measuring PUE? PUE is the ratio of "power in at the utility meter" to "power that actually makes it to the IT equipment."[8] A high ratio indicates an inefficient facility; much of the incoming power is being used up by the cooling system and/or lost in conversion within the UPS system, rather than making it to the useful IT equipment. If a manager's goal is to determine which facility is less efficient overall—perhaps so that one lease can be terminated and a better facility found—then these ratios might be enough information. If the goal is to improve the efficiency of the facilities by replacing equipment, this is not enough information to make conclusions about whether the energy waste results from an inefficient UPS, poor airflow, a worn-out chiller, or something else.

Understanding the limits of data is even more important where data center reliability is concerned. For example, many building management systems (BMS) receive data periodically on "the last known good state" of various pieces of equipment. However, checking a log of these readings does *not* enable managers to conclude that the data center is functioning normally. When a critical piece of equipment breaks, its "last known good state" is (by definition) that it is working. In one data center, a BMS failed to detect an equipment failure for precisely this reason. Luckily, the data center in question had an additional monitoring scheme, which detected the problem and alerted operators.[9]

Normalize measurements to account for external factors

Data center managers should never forget to take the simple measurements that are necessary to put other operating characteristics into context. For example, managers should always account for growth (or decline) in IT equipment loads when trying to evaluate the impact of efficiency initiatives.[10] Even if an efficiency initiative is quite successful,

[8] PUE stands for Power Usage Effectiveness, and is a ratio defined as (Total facility power) / (IT equipment power). See Belady 2007.

[9] See *Case Study: Availability Improvements Using Granular Monitoring at Sybase*, p.4-7.

[10] Growth in IT equipment can mean many things: growth in number of servers, aggregate IT power demand, total compute cycles provided, number of users served, etc. Data center managers may wish to track several of these variables to place other data in context.

data center energy use may continue to increase year after year if new servers are constantly being added.[11]

Weather and climate are also important factors for which to control. Since HVAC systems operate more efficiently (and use their economizers more) when it is cooler, taking careful note of weather allows managers to correctly compare the performance of facility cooling systems in, for instance, their New York and Florida locations. Even at a single location, changes in weather from month to month might mask important cooling system trends if not handled carefully. (Weather data is a good example of data that might not be needed immediately, but should be collected for potential use in the future.)

It is especially important for managers to take careful notes if they plan to make multiple changes to a data center at the same time. In one corporate data center, managers found it challenging to determine the energy savings from improving computer room air flow over a period of time, because the data center had been increasing the use of its economizer during that same period. The data center had excellent records of energy use by the chiller, but virtually no records of which days the economizer was running. It was unclear how much of the energy savings should have been attributed to the air flow improvements and how much was just a result of increased economizer use.[12]

Measure what's important, not just what's easy to measure

Data center managers should resist the temptation to measure what is easy rather than what is important. For example, in many cases, a data center will be only one small part of a larger office building, and typically there is only one utility meter measuring electricity consumption for the entire building. Managers can easily get aggregate building electricity use, but meaningful information about the data center itself will be obscured. Managers who are serious about monitoring might consider installing a separate electricity sub-meter for the data center or taking other measurements to help "screen out" the effects of the non-data center spaces.

Another difficult thing to measure in a data center is "IT productivity." How much business value does a data center end-user get out of the servers, storage, and network devices that are deployed? Some data center owners measure cost reductions by tracking the total cost of ownership (TCO) per server deployed. However, what if Server A has double the TCO of Server B, but Server A does three times as much useful computing? Server A is still a better deal. However, "useful computing" is tough to measure, and the definition changes depending on whether the server is hosting web pages, delivering emails, or processing database transactions. One might measure a server's number of calculations per second, but what if a competing product runs better-designed software that can perform more business transactions with fewer calculations?[13] IT productivity is an ambiguous concept, but in many cases it is the measure of interest.

In any case, data center managers should do their best to track what is important to their business, while recognizing that it will indeed be impossible to directly measure some very important things.

---

[11] See *Case Study: Cooling & Chilled Water Efficiency Project at Sybase*, 17-19, for an example in which server growth partially masked the energy savings generated by changes to a data center's cooling system.

[12] See *Case Study: Cooling & Chilled Water Efficiency Project at Sybase* , p.17.

[13] See Stanley et al. 2007 for a discussion of some of the key aspects of IT productivity by the Uptime Institute. EPA, the Green Grid, and others are also continuing research in this area.

After measuring, take action

After any period of monitoring and collecting data on their facilities, managers should remember to *take action* based on what they have learned. The goal of monitoring is to collect information to inform decisions so that *action* can be taken to further business objectives. After taking action, managers should continue collecting data in order to receive *feedback* about whether or not the actions were effective. This feedback often forms the basis for new goals and/or actions.

Figure 2 below shows the cyclical pattern of goals, action, and feedback.[14] **Figure 3** below uses a (fictional) example of a data center improvement to illustrate the Cycle of Action and how it works with data to enable a manager to pursue business objectives. This cyclical process can be used to design, execute, evaluate, and (if necessary) refine a "one shot" improvement project, but it also lends itself quite well to initiatives focused on continuous, incremental improvement.

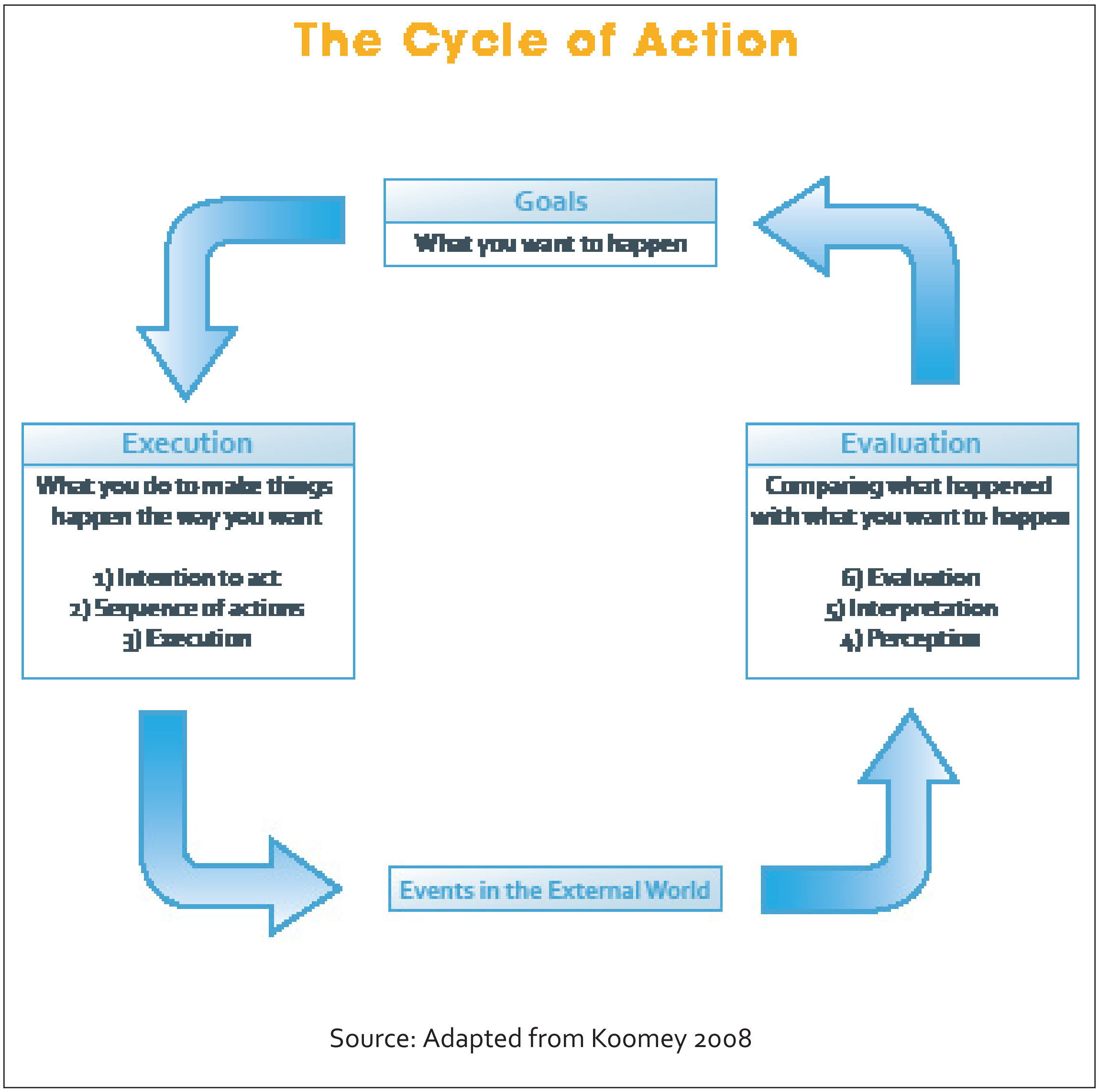

Figure 2 – *The Cycle of Action*

---

[14] The figure is adapted from Koomey 2008 *Turning Numbers into Knowledge* (p.10-15), which in turn adapts the figure from *The Design of Everyday Things*, by Donald Norman (1990).

Example: Cycle of Action for Cost Reduction in a Data Center		
Stage	**Use data to…**	**Example insight**
Goals	*Determine and prioritize goals.*	No availability problems in two years, so let's focus now on reducing TCO. The current TCO "baseline" is X $/kW-month.
Execution 1) Develop an intention to act		Let's cut TCO by 5% this year.
2) Formulating the action sequence	*Determine the action sequence*	Cooling system data shows a poorly performing chiller driving up electricity costs. That's what we need to fix.
3) Executing the action sequence		(Make changes to chiller)
Events in the external world		(Operate for another year and collect data.)
Evaluation 4) Perceiving the state of the external world	*See the results*	I see that chiller energy use dropped 15% vs. last year.
5) Interpreting the perception	*Figure out what the results mean*	Weather data confirm that the energy reduction is "real" and not just a result of a cool summer. In fact, our % savings is actually a bit higher than it looks because we added a lot of servers this year, offsetting some of the savings.
6) Evaluating the interpretation and comparing with your goals	*Evaluate your actions*	The changes to the chiller cut TCO by 7%. We beat our goal!
Goals (the cycle continues)	*Determine and prioritize goals*	We could cut TCO more, but we're seeing a trend of server growth. Let's focus on capacity planning this year…

Figure 3 – *The Cycle of Action, illustrated with a fictional example.*

How to measure – Details of deploying an effective measurement system

After a data center manager has carefully thought out his or her measurement needs, he or she can begin to consider the capabilities of different monitoring systems. Some high-level considerations are listed below.

Key considerations:

- Ability to collect data from all desired devices
- Granularity of data collection
- User friendliness and ease of integrating data across devices and time scales
- Scalability for mass deployment and multi-site capability
- Adaptability to new measurement needs
- Trending and analysis of data
- Integration with control systems
- Ability to detect problems and notify data center operators
- Specifications of sensors – Resolution and drift

Ability to collect data from all desired devices

All data centers include different types of equipment—servers, networking hubs, cooling systems, power distribution units, temperature sensors, smoke alarms, etc. Even within a single equipment category such as Computer Room Air Conditioning (CRAC) units[15], a data center may have over time acquired several different models or brands. So, a good data center measurement system should be able to easily collect data from all the devices that managers must monitor in order to meet their business objectives of higher productivity, lower cost, and improved uptime.

For example, imagine a data center manager attempting to measure the ratio of "power in at the utility meter" to "power that actually makes it to the IT equipment." Even in the simplest case, measuring this ratio would require communication with two pieces of equipment. The building's (intelligent) utility meter must provide data about incoming power, and the facility's UPS system must report the amount of output power it is providing to the IT loads. A more detailed breakdown of energy use (perhaps as part of an efficiency initiative) would require communication among even more devices—perhaps separate monitoring of cooling, ventilation, and lighting systems.

In addition to pulling data from larger pieces of "smart" equipment[16] such as UPS and CRAC units, an ideal measurement system should also be able to communicate with smaller external sensors that managers might place on an ad-hoc basis. For example, an external temperature sensor might be used to cross-check temperature readings from a nearby CRAC unit or to take measurements at a specific hot spot of interest.

Many data center equipment vendors have monitoring systems that can pull data from their own equipment, but many of these cannot easily communicate with competing vendors' equipment. An ideal monitoring system would be vendor-neutral, able to pull data from any device type, brand, or model.

[15] CRAC units are modular cooling devices, which can be deployed around a data center.
[16] *i.e.*, equipment that has the built-in capability to collect data on its own operation or relevant external variables and to communicate these data via some protocol. Many pieces of data center equipment available from major vendors have such abilities.

Granularity of data collection

Granularity of data collection enables managers to really understand their data center's operation—detecting areas for improvement, pinpointing problems, and specifically measuring the effectiveness of corrective actions. A good measurement system provides several types of granularity. **Figure 4** below summarizes these types.

Types of granularity

Type	Example levels	Sample data center issue
Space	Temperature measurements aisle-by-aisle or rack-by-rack. Current measurements circuit-by-circuit.	*Where are the hot spots? Are any power distribution branches close to being overloaded?*
Time	Data points each month, day, hour, or minute	*Are afternoon cooling loads much higher than the daily average? Does server power draw ever "spike?"*
Across devices	Groups of CRAC units vs. individual CRAC units	*Which CRAC units are overworked?*
Within a single device	Current on each of the three power phases in a single Power Distribution Unit (PDU)	*Is there a current imbalance among phases that could compromise reliability?*

Figure 4 – *Types of granularity*

Granularity in **space** makes data useful for pinpointing location-based data center problems, such as hot spots or overloaded branch circuits.

Granularity in **time** is important when data need to be analyzed for relatively short term phenomena. At one data center, current measurements at intervals of around 10-15 minutes serendipitously discovered that a set of backup servers had been programmed to run their backup routines in the middle of the afternoon—just when cooling loads were already high and electricity was at its most expensive. After this was discovered, it was easy to reprogram the routines to run in the middle of the night. Time granularity was critical, since daily average measurements would not have discovered this problem.

Granularity **across devices** allows operators to see trends at the level of individual devices rather than simply viewing aggregate information from groups of devices. At the same data center mentioned above, tracking of individual CRAC units discovered that some were overloaded while others did not need to provide much cooling at all. Seeing data at this level helped the data center managers identify and eliminate hot spots.[17] For mixed use-buildings, tracking multiple pieces of equipment separately can also help managers disaggregate the energy performance of the data center space from that of the rest of the building.

[17] See *Case Study: Cooling & Chilled Water Efficiency Project at Sybase.*

Granularity **within a single device** can identify other problems. Separately monitoring each of the three phases of power within a Power Distribution Unit (PDU) can help identify reliability-reducing current imbalances. Multiple parameters within a CRAC unit can be used to detect airflow problems or fan-belt slippages.[18] [19]

In general, data center managers should take measurements at fine enough granularity to meet their data needs, and they should rely on the measurement system to aggregate data when needed.

User friendliness and ease of integrating data across devices and time scales

A data center monitoring system must be as user friendly as possible. Even the best data is worthless if managers do not use it, and managers will be much more likely to use it if using it is easy. A manager should be able to conveniently add, remove, or change the data logging devices that are part of the monitoring scheme. Also, although the monitoring system might collect data from diverse equipment at varying levels of granularity, it must integrate this information for easy viewing and analysis.

For example, consider an initiative to reduce data center TCO by improving cooling system efficiency, in which the first step is to investigate the cooling system's sensitivity to weather conditions. Even if a data center's chiller and its ambient weather sensors would not normally deliver data in compatible formats, the monitoring system should bridge the gap and allow a manager to easily plot data on chiller energy use against weather data from the same period.

Integrating data across different time scales should also be easy. At any given time, a data center manager may wish to view annual data for tracking long term trends or minute-by-minute data for troubleshooting specific problems. A manager will also need to convert data time scales to facilitate comparisons. For example, in one data center that was evaluating the impact of cooling efficiency improvements, it was useful to convert hourly data on chiller power use into monthly averages so that it could be plotted against monthly estimates of server load.[20]

Scalability for mass deployment and multi-site capability

A data center monitoring solution should also be scalable. Ideally, a system would be appropriate for a small data center or test deployment as well as being able to handle the demands of an enterprise-class facility. The system should also be able to scale across multiple data center locations, enabling knowledge sharing among facilities and easier corporate oversight of the IT and data center functions. Finally, the ability to integrate other locations can allow managers to monitor smaller facilities (such as server closets and server rooms) that might not be worth monitoring on their own, improving reliability and efficiency at these sites as well.[21]

A good measurement system should also be able to scale up without shutting down. For example, once the system is deployed in one part of a data center, managers should be able to add another server rack, another IT room, or another site to the monitoring scheme without interrupting data collection from the original installation.

---

[18] For example, if a CRAC unit is using a lot of chilled water but has a low measured return air temperature, then data center operators should be suspicious of a problem. High chilled water use means the unit is *trying* to provide a lot of cooling, but the low return temperature means that the cold air being blown has not succeeded in pulling much heat away from the IT equipment.

[19] See *Case Study: Cooling & Chilled Water Efficiency Project at Sybase*, p.14-15, and *Case Study: Power Chain Capacity Expansion at Sybase* for more information on detecting current imbalances and CRAC unit problems by monitoring multiple parameters within a single device.

[20] See *Case Study: Cooling & Chilled Water Efficiency Project at Sybase*, p.17-19.

[21] These spaces are not insignificant. In 2006, server rooms and server closets accounted for just under 30% of total data center electricity use in the U.S. (EPA 2007, p.27).

Adaptability to new measurement needs

Monitoring solutions must also be adaptable, since it is impossible to anticipate all the needs of data center managers in the future. The system should be flexible enough to integrate new types of equipment, so that data can be collected from *next* year's UPS addition as well as this year's installed model. Such adaptability will be even more important in coming years as monitoring and energy management extend to IT equipment itself, instead of being confined to facility infrastructure equipment for power, cooling, etc.

For example, in future data centers, an advanced cooling system reaching its limits may be able to instruct non-essential servers to temporarily ramp down processing speed and generate less waste heat. There are already systems available for enterprise tracking of IT assets, and it better integration of IT and facilities monitoring will likely be desirable to continue increasing data center productivity. A good monitoring solution should be able to accommodate such future developments.

A monitoring system should be adaptable in its output as well as its data collection. Data center managers may need to run different statistical reports or track different facility metrics, or they may need to provide the monitoring system's real-time or historical data to new third-party software in future years.

Some monitoring systems already incorporate adaptability features. Modular hardware and software design in a monitoring system can allow vendors or third parties to create "plug and play" modules, which allow a new piece of equipment to communicate with the monitoring system. Also, a monitoring system might allow managers to define and track their own metrics of data center performance, derived from the raw data collected by monitoring devices.

Trending and analysis of data

In addition to collecting data, successful measurement systems must be able to store, trend, and analyze data. Trending is critical, because it allows managers to see how a data center's performance characteristics change over time. For example, CRAC units often cycle up and down over the course of a day in response to changing cooling demand. So, a single reading can be hard to interpret; the observer does not know if it represents a high or low point in operation. (As one experienced data center manager put it, "Snapshots are worthless.") Multiple readings forming a trend allow managers to really understand their facility.

Trending also enables managers to take action and make improvements. Since reliability is critical, data center managers (and their bosses) are often wary of making changes based on spot measurements. However, patterns of historical data showing clear trends can enable managers to make decisions with confidence.

Integration with control systems

In addition to presenting information to managers so that they can take action in the data center, a good monitoring system should be able to integrate with control systems so that action can be taken *automatically* when desired. Obviously, many building management systems and pieces of data center site infrastructure can do some of this already—*e.g.*, systems can already instruct generators to activate when they detect a power grid outage.

A good monitoring system will have controls integration that keeps pace with the level of data collection detail that can be provided. For example, a monitoring system that can collect rack-by-rack temperature measurements in real time should be able to feed this information to dynamic cooling optimization controls.

Ability to detect problems and notify data center operators

A monitoring system must be able to both (1) detect problems in the data center and (2) notify data center operators so that they can take corrective action before an outage occurs. A system that cannot do both of these things does not do much to improve reliability.

A good way for a monitoring system to detect problems is to *actively poll* the devices it is monitoring. If the monitoring system ever fails to receive a timely response from a device, it knows there is a problem. Note that many building management systems simply wait passively for subordinate devices to send updates about their operation. But if one of these devices freezes up, it will never send a message saying that it is broken, and the building management system will never know.

After detecting a problem, a monitoring system should be able to actively send an alarm (via email, pager, etc.) to appropriate personnel. This type of feature is common in almost any system used to monitor data center operations.

Finally, the data collection system *itself* should be reliable and possibly able to accommodate some type of redundancy if parts of it fail. Most importantly, the failure modes of the monitoring system must be *different* from the failure modes of the equipment being monitored. For example, if a chiller and the monitoring system fail at the same time because they are on the same circuit, then there is no cooling system *and* no warning of a problem.

In many data center facilities, redundancy can be achieved simply by combining a building management system (with control capabilities) and an additional measurement system (with better measurement and trending capabilities). Although the two systems do not have 100% overlap in their functionality, they can back each other up to some extent. For example, one data center falls back on its measurement system to provide status information when the BMS temporarily goes off line.[22]

Specifications of sensors – Resolution and drift

Finally, any data collection system begins with sensors that actually perform the measurements of temperature, airflow, IT load, etc. Often, data center managers will not need to consider sensors directly, since many types of data center equipment already come with their own sensors installed. For example, such sensors are what enable "smart" CRAC units to report their inlet air temperatures or chilled water use. Data center managers should consider at least two things about sensors:

<u>What is the resolution of the measurements?</u>

> For example, imagine that temperature Sensor A is accurate to +/- 1 ºF and Sensor B is accurate to +/- 0.01 ºF. When Sensor A reads 50 ºF, it means that the true temperature is between 49 and 51 ºF. When Sensor B reads 50.00 ºF, it means that the true temperature is between 49.99 ºF and 50.01 ºF. Some data center variables require more exact measurements than others.[23]

<u>How often do the sensors need to be recalibrated?</u>

> Sensor "drift" refers to the process of sensors losing accuracy over time. Humidity sensors often have problems with drift and thus need to be recalibrated frequently.

---

[22] See *Case Study: Availability Improvements Using Granular Monitoring at Sybase*, p. 9-10.

[23] For example, a measurement error of +/- 1 ºF is ok when tracking air temperatures, since it does not matter much if the computer room is at 75 or 76 ºF. However, when measuring the waterside temperature difference across a typical chiller (which might only be a few degrees) a pair of temperature sensors with an error of +/- 1 ºF would be almost useless (Greenberg et. al. 2006, p.7). Imagine that the sensor on one side of the chiller reads 50 ºF and the sensor on the other side reads 45 ºF, for an apparent calculated difference of 5 ºF. But since each sensor could be off by a degree, the actual difference in temperatures could be as little as 49 – 46 = 3 ºF or as large as 51 – 44 = 7 ºF. That's a spread of more than a factor of two!

Conclusions

Effective measurement and monitoring can provide data to inform decisions that enhance the business value of a data center—increasing productivity, reducing costs, or improving reliability. Furthermore, there are opportunities to improve efficiency through a continuous process of incremental changes, as well as through the more traditional "one shot" project-based approach.

When considering possible monitoring strategies or tools, managers should keep in mind their specific business objectives for the monitoring effort. They should think carefully about their data needs *before* deciding on a particular tool or approach, thinking about the limitations of the proposed data to be collected and what external factors may also need to be tracked in order to properly interpret results. It is also important to remember to take action based on the data they have collected, and they should continue collecting data to provide feedback on the effectiveness of these actions. Finally, managers should consider many strategic and technical aspects of potential monitoring solutions. The ability to collect data from all desired devices, granularity of data collection, user friendliness, scalability, and adaptability are key. Trending and analysis capabilities, integration with control systems, and the ability to detect problems in the data center are also important.

Finally, it is worth noting that there is a wide variety of hardware and software products available for data center monitoring. Managers should think carefully about the capabilities they need when selecting a particular product or vendor.

References

Belady, Christian (editor). 23 Oct 2007. *Green Grid Data Center Power Efficiency Metrics: PUE and DCIE*. Beaverton, OR: The Green Grid. White paper # 6, Rev 2008-0. Available at <http://www.thegreengrid.org/~/media/WhitePapers/White_Paper_6_-_PUE_and_DCiE_Eff_Metrics_30_December_2008.ashx?lang=en>. Accessed 21 Jul 2009.

Environmental Protection Agency (EPA). 2 Aug 2007. *Report to Congress on Server and Data Center Efficiency*. EPA. Available at <http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf>. Accessed 14 Jan 2009.

Greenberg, Steve, William Tschudi, and John Weale. May 2006. *Self Benchmarking Guide for Data Center Energy Performance*. Berkeley, CA: Lawrence Berkeley National Laboratory (LBNL). Available at <http://hightech.lbl.gov/documents/DATA_CENTERS/self_benchmarking_guide-2.pdf>. Accessed 7 Sep 2009.

Koomey, Jonathan. 2008. *Turning Numbers into Knowledge: Mastering the Art of Problem Solving*. 2nd edition. Oakland, CA: Analytics Press. <http://www.analyticspress.com/>.

Koomey, Jonathan, et al. 2007. *A Simple Model for Determining True Total Cost of Ownership for Data Cent*ers. Santa Fe, NM: Uptime Institute. Available at <http://uptimeinstitute.org/content/view/21/55/>. Accessed 19 Sep 2009. (The original white paper was released in 2007, and an updated version 2.1 of the TCO spreadsheet model was made available as of 31 March 2008.)

Kumar, Rakesh. 2 Mar 2009. *Continuously Optimize Your Data Center Capacity Before Building or Buying More*. Stamford, CT: Gartner, Inc.. Available at <http://www.gartner.com/DisplayDocument?doc_cd=165501>. Accessed 7 Sep 2009.

Modius, Inc. Nov 2008. *Case Study: Availability Improvements Using Granular Monitoring at Sybase*. San Francisco, CA: Modius, Inc.. Available at <http://www.modius.com/resources>. Accessed 7 Sep 2009.

Modius, Inc. Nov 2008. *Case Study: Cooling & Chilled Water Efficiency Project at Sybase*. San Francisco, CA: Modius, Inc. Available at <http://www.modius.com/resources>. Accessed 7 Sep 2009.

Modius, Inc. Nov 2008. *Case Study: Power Chain Capacity Expansion at Sybase*. San Francisco, CA: Modius, Inc. Available at <http://www.modius.com/resources>. Accessed 7 Sep 2009.

Norman, Donald A. 1990. *The Design of Everyday Things*. New York, NY: Doubleday/Currency.

Stanley, John, Ken Brill, and Jonathan Koomey. 2007. *Four Metrics Define Data Center "Greenness."* Santa Fe, NM: Uptime Institute. White paper TUI3009F. Available at <http://uptimeinstitute.org/>. Accessed 8 Jan 2009.

About the authors

John Stanley currently works at a firm specializing in data center energy efficiency research and consulting. He holds a Master's degree from the Energy and Resources Group at the University of California, Berkeley, where his thesis research focused on energy efficiency in data centers. While a graduate student, he worked at the Uptime Institute with founder Ken Brill on publications including *Four Metrics Define Data Center "Greenness."* He also worked at Lawrence Berkeley National Laboratory as a contributing author to the EPA's 2007 *Report to Congress on Server and Data Center Energy Efficiency*. Prior to graduate school, John worked at Rocky Mountain Institute as part of its energy efficiency consulting practice.

Jonathan Koomey is a Project Scientist at Lawrence Berkeley National Laboratory and a Consulting Professor at Stanford University. Dr. Koomey is currently (Autumn 2009) a visiting professor at Yale University's School of Forestry and Environmental Studies. Dr. Koomey is one of the leading international experts on electricity used by computers, office equipment, and data centers, and is the author or co-author of eight books and more than one hundred and fifty articles and reports on energy and environmental economics, technology, forecasting, and policy. He has also published extensively on critical thinking skills. He holds M.S. and Ph.D. degrees from the Energy and Resources Group at the University of California at Berkeley, and an A.B. in History of Science from Harvard University. In 1993 he won the Fred Burgraff Award for Excellence in Transportation Research from the National Research Council's Transportation Research Board. He was named an Aldo Leopold Leadership Fellow in 2004 and an AT&T Industrial Ecology Fellow in January 2005. He has been quoted in the New York Times, the Wall Street Journal, Barron's, The Financial Times, The Washington Post, Science, Science News, American Scientist, Dow Jones News Wires, and the Christian Science Monitor, and has appeared on Nova/Frontline, BBC radio, CNBC, All Things Considered, Marketplace, On the Media, Tech Nation, and the California Report, among others. His latest solo book is *Turning Numbers into Knowledge: Mastering the Art of Problem Solving* <http://www.analyticspress.com/>, now out in its second edition (April 2008). For more biographical details and a complete publications list, go to <http://www.koomey.com/>.

Acknowledgements

This report was produced with the generous cooperation of Modius, Inc. <http://www.modius.com/>. Many of the concepts and anecdotes in this report draw from three case studies in the *Measurement Series*, published by Modius. In addition, Modius provided helpful feedback on drafts, support for document layout and production, and assistance in publicizing this report. In particular, the authors would like to thank Craig Compiano, Donald Klein, and Earl Sacerdoti.

All errors and omissions in this report are the responsibility of the authors alone.

Copyright notice

Suggested citation:

John Stanley and Jonathan Koomey. 2009. *The Science of Measurement: Improving Data Center Performance with Continuous Monitoring and Measurement of Site Infrastructure*. Oakland, CA: Analytics Press. October 23. <http://www.analyticspress.com/scienceofmeasurement.html>

Appendix: Summary table of references to other Measurement Series documents

For readers interested in the details of the *Measurement Series* case studies, **Table A-1** below summarizes the references to these case studies made in this report.

Summary of Measurement Series References in This Report

Insight	Mentioned in this doc, page...	More information in case study...	Case study page
Iterative approaches and incremental improvements over time can add up. One data center was able to realize significant savings by inching its chilled water set point up by 9 °F over a period of about two years.	2	Cooling	6-8,17
Power use trend data helps build a case to management for a facility upgrade	4	Case Study: **Power** Chain Capacity Expansion at Sybase	12
A BMS fails to notice a facility failure with its passive data collection, but a monitoring system with "active polling" of devices detects the problem and alerts operators.	5	Case Study: **Availability** Improvements Using Granular Monitoring at Sybase	4-7
It is important to account for growth (or decline) in IT load over time when evaluating the results of efficiency initiatives.	6	Case Study: **Cooling** & Chilled Water Efficiency Project at Sybase	17-19
It is important to take careful notes if multiple facility changes are happening at the same time. In one corporate data center, it was difficult to determine the energy savings directly attributable to airflow improvements, because economizer use was increasing as well.	6	Cooling	17
Granularity in monitoring across devices is useful. One data center tracked individual CRAC units to hunt down hot spots.	10	Cooling	p.7, whole doc
Granularity in monitoring within a single device is useful. Examining multiple variables within a single CRAC unit can help find problems like air-flow "short circuits" or fan-belt slippage.	11	Cooling	14-15
Examining multiple variables in a PDU can detect current imbalances.	11	Power	p.7, whole doc
A good monitoring system makes it easy to aggregate data at different time scales. One data center condensed hourly chiller data into monthly averages to compare against monthly numbers for server load, as part of a cooling savings evaluation.	11	Cooling	17-19
Redundancy in monitoring is useful. During a temporary BMS outage, one data center used its monitoring system to keep watch for problems.	13	Availability	9-10

Table A-1 – *Summary of Measurement Series references in this document*