



The Data Center Observatory (DCO) is a centerpiece of Carnegie Mellon's attack on ever-growing data center operational costs. As a data center, it will provide a computation and storage utility to resource-hungry research activities such as data mining, design simulation, network intrusion detection, and visualization. As an observatory, it will provide invaluable real data to systems researchers seeking to understand the sources of operational costs and to evaluate novel solutions. Combining the two builds on Carnegie Mellon's tradition of actively using and show-casing new computing approaches, even as we invent them, allowing us to push the frontiers and stay at the forefront of technology.

WHY IS THE DCO NEEDED?

Operational costs are spiraling out of control. Power and cooling now costs as much computer acquisition, and human administration costs are 4–7X higher than capital expenditures. Worse, both of those costs grow with every generation of computing equipment—for example, the power demand of a computer grows with its speed. Worse still, many of the costs are poorly understood—in particular, it remains something of a mystery where all of that human administration time goes, which makes it difficult to address well with new technologies.

HOW DOES THE DCO HELP?

The DCO enables us to target and attack the real challenges of data centers. First, it provides us a live environment to analyze, exposing where the time goes and where the costs are. The DCO has been designed, from the beginning, with detailed instrumentation at every level, from the power/cooling to the software systems to the human administration time. Researchers will be able to create detailed breakdowns over the long-term and at particular points in time (e.g., during failure events), as well as drive automated problem detection and diagnosis research. Second, it provides us a real environment in which to test reasonably mature new technologies and measure how well they work. Without the DCO, researchers are left with little evidence to corroborate new ideas beyond argumentation, leading many to work on other problems. The DCO allows us to tackle one of the largest IT challenges of our time: data center operational costs.

EARLY RESEARCH ACTIVITIES

The DCO enables a broad range of research activities, beyond “simply” measuring and understanding operational costs. A few initial thrusts include: (1) *Adaptive power management*. Carnegie Mellon teamed with APC to take advantage of their novel hot air containment approach, achieving energy savings from the beginning. We are also teaming with APC to explore new approaches to dynamically controlling which computers are on and off, based on application demands, saving energy at every level of the system. (2) *Automated storage management*. Carnegie Mellon's Parallel Data Lab (PDL) has been developing new architectures and tools for mitigating storage administration costs, and the Self-* Storage system that they are building will provide the DCO's storage. (3) *Finger pointing in large-scale systems*. Diagnosing problems is a notoriously difficult problem in data center environments, and we are combining the DCO's detailed instrumentation with on-line and off-line machine learning tools to automate the process of identifying the root causes of problems that arise.



SOME DETAILS

- The DCO is located on the main floor of the Collaborative Innovation Center (CIC).
- The location was chosen to maximize visibility, and has a windowed wall for walk-by tours and an LCD display to continuously tell the research story and show measurement data (thermal graphs, utilizations, etc.).
- The DCO is 2,000 square feet in size, with an adjacent 1,100 square foot lab for administrator cubicles, storage, and unpacking and testing equipment.
- The DCO has been designed, in partnership with APC Corporation, to be fully instrumented and adaptively controlled, allowing measurement of all of the environmental aspects, including power consumption, temperature, humidity, chilled water consumption rate, and fan speed.
- APC In-Row™ cooling and Hot Aisle Containment™ enable remarkable equipment densities; the room has a peak capacity of 40 racks of computing equipment that consume 774kw of power.
- In the event of a loss of power and/or the loss of the campus chilled water supply, the DCO has electrical and mechanical provisions to facilitate a controlled 4-minute shutdown.
- The Grey System, a universal and highly-secure phone-based access control system developed by Carnegie Mellon CyLab researchers, is used to control access to the DCO.
- Fiber optic network connects the DCO to the primary campus backbone.
- Construction was completed in April, 2006, and computing equipment went on-line immediately thereafter.

PARTICIPANTS

The DCO is a huge, long-term, collaborative endeavor that includes many organizations across Carnegie Mellon and elsewhere. Carnegie Mellon participants include: the Carnegie Institute of Technology College of Engineering, the Department of Electrical & Computer Engineering, CyLab, the School of Computer Science, Facilities Management Services and the Parallel Data Lab. Industry partners include: APC, EMC, EqualLogic, HP, Hitachi, IBM, Intel, Microsoft, Network Appliance, Oracle, Panasas, Seagate, Sun and Symantec. Government sponsors include: the Air Force Office of Scientific Research, the Army Research Office, the Defense Advanced Research Projects Agency, and the National Science Foundation.

FOR FURTHER INFORMATION

Additional information is available at <http://www.pdl.cmu.edu/DCO/>. Also, you may contact us directly at:

Prof. Greg Ganger
Director, Parallel Data Lab
ganger@ece.cmu.edu

Bill Courtright, Ph.D.
Executive Director, Parallel Data Lab
wcourtright@cmu.edu

