WHITE PAPER

Written By: Paul Blevins, CFO; Greg Maynard, CTO **Release Date: 12-18-09**

Written For: SDG&E Bid Program

Subject: Onsite Customer Data Operations Energy Efficiency Project

The following was written and submitted to San Diego Gas & Electric as part of an application for energy efficiency cash incentives to be paid to Onsite's customer by the local utility in connection with a major re-structuring of their IT infrastructure and operational process. It is estimated that the energy incentives from the program will exceed $2 million and that annual energy savings will exceed $2.4 million per year. The term "ONSITE CUSTOMER" is used in place of the actual company on who's behalf Onsite Energy submitted the application. Reprints of this white paper can be obtained by contacting:

Barbara Schoor
Onsite Energy Corporation
2701 Loker Ave. West, Suite 107
Carlsbad, CA 92010
Phone: 760-931-2400
Email: bschoor@onsitenergy.com

ONSITE CUSTOMER QRC HIGH PERFORMANCE DATA CENTER PROJECT

The IT Department at Onsite Customer in San Diego desires to implement comprehensive, enterprise-wide changes in the way they serve their user base. These comprehensive changes will dramatically improve the energy efficiency of their enterprise-wide data center operations resulting in millions of kilowatt hours saved over the coming years.

Because of Onsite Customer's business needs, it is heavily reliant on a high degree of data center support. The most compute intensive IT processes are associated with product development activities that include integrated circuit design, firmware development and software development. Other compute intensive processes include the virtualization of enterprise application environments in addition to other office productivity services . Most of these applications are "Mission Critical" and operate 24/7 to support on-going operations.

Onsite Customer currently employs over 16,000 people in 37 separate facilities. As Onsite Customer grew and expanded, it needed to add IT capacity very quickly to support its operations. This IT capacity grew organically and was specifically tailored to specifications required by the operating division or department the capacity was intended to serve. For instance, the chip design group wanted their IT system, finance folks wanted theirs, the customer support department wanted theirs, etc. The growth pattern was chaotic as the IT department at Onsite Customer worked to keep up with the frantic pace of growth at Onsite Customer. The current geographical dispersion of IT assets throughout the enterprise reflects this growth pattern. Network latency over internet and/or WAN connections also limited the ability of IT department to centralize their support assets and thus fostered decentralized growth.

In addition to geographic inefficiencies, inefficiencies developed in software topography embedded on physical IT assets (servers, drives, routers etc). This software consists of applications used by the Company in addition to operating systems and data. In general, the software is very thinly spread over a large number of physical servers that were specified at the time the software was acquired for use. At the time of installation, functionality was foremost and very little attention was given to relatively low utilization of the physical server's compute capacity. As a result, inefficiencies in the form of underutilized server capacity developed. Onsite Customer's management estimates that on an enterprise-wide basis, the average CPU utilization is about 5%. From an energy perspective, this means that on average, 95% of the time a typical server is on, but it is not performing any useful task.

Onsite Customer's Solution

The Management at Onsite Customer recognize that major improvements to these inefficiencies can be made by implementing enterprise-wide, comprehensive (or holistic) changes in the IT infrastructure of the Company.

The core of Onsite Customer's strategy is to restructure the IT asset base by transforming it from a decentralized collection of relatively inefficient data centers to a highly centralized and efficient data center operation. To achieve this, Onsite Customer wants to implement IaaS (Infrastructure as a Service) based operating environments. These environments enable the IT management to locate and configure physical IT assets in a highly efficient centralized location while at the same time creating virtual IT asset structures that can serve the user base with the same flexibility as the decentralized IT asset operating model once did.

Onsite Customer plans to implement this transformation using the following strategies and technologies:

1. **Identify optimal location for central data center** – Onsite Customer's main campus located San Diego was selected because of its proximity to the Company's combined heat and power ("CHP") system, the proximity to the facility's very efficient chilled water plant ("CWP") (with

back-up cooling), the proximity to Company Headquarters, and the existence of one of the Company's largest conventional data centers at the location.

2. **Upgrade physical servers to high performance state-of-the-art technology** – In order to run virtualization/loadshare platforms and to provide satisfactory response times under an IaaS environment, very powerful blade servers must be purchased. The servers will be quad core/dual processor, 2.5 to 3.0 GHz and be specifically designed to support the virtualization of the hardware platform layer running multiple simultaneous applications or jobs distributed over clusters of physical servers. Compared to old technology blade servers, these servers are designed to serve multiple simultaneous user loads or application evnironments and operate at average CPU utilization levels of 50% or more.

3. Implementation of virtualization layers into the data center – Installation of virtualization layers will enable IT management to separate the physical configuration of the servers from a virtual or "apparent" server configuration as required by the various users throughout the company. This means IT management can optimize physical data center asset operation while retaining the flexibility needed to meet the needs of the user base. Virtualization software enables near-dynamic provisioning of compute capacities through abstraction of physical hardware, combining the aggregated demand of all users to run across multiple consolidated physical servers. This is achieved by enabling a cluster of servers to support multiple simultaneous applications and processes. The result is that fewer physical servers are required to support a given number of applications and users. The physical servers in the cluster actually share and allocate their compute capacity based on compute demand, thereby increasing the number of applications and users per server. This sharing capability enables the servers to operate at 50% or higher CPU utilization compared to 5%, which is Onsite Customer's current estimated utilization.

4. Migrate to containerized data center support structure – The increase in compute capacities of the physical servers/utilization by IaaS create higher concentrations of heat per given unit of rack space in the data center. To be able to remove this heat and transfer it out of the data center space (to the chilled water loop in Onsite Customer's case), will require an extremely efficient transfer environment. The current conventional data center at Onsite Customer cannot effectively achieve this. Therefore Onsite Customer plans migrate to a containerized platform for storage and operation of the high performance servers. The container is a self-contained enclosure specifically designed to circulate air to transfer heat from the physical servers to heat exchangers which transfer the heat into the chilled water loop. The containerized storage approach uses about 10% of the amount of energy required by conventional data centers to transfer this heat. The container achieves these efficiencies through virtual elimination of hot/cold air mixing, CRAC conflicts, air turbulence and air travel distances.

5. Decommission/limit growth of disparate data center operations – Once the container, servers and virtualization platform is installed, the IT group will begin the process of migrating applications and user load from the company's existing disparate and inefficient physical platforms to the new efficient one. This will be achieved by establishing a virtual replica of the old data center and migrating the applications onto it thus achieving total transparency for the user base. Once the migration is complete, the old data center will be decommissioned.

The execution of these 5 steps will result in a rapid increase in the utilization of the new data center. In addition, natural growth in data processing demand will further increase the utilization of the new data center.

Implementation Project

The implementation of the IaaS platform will consist of the following phases:

1. QRC Data Center Expansion – This includes installation of the container, high performance servers and virtualization layer (across all suitable physical servers in the expanded QRC data center). ETC- March 2010

2. Transfer of Services to QRC Virtual Platform – This will include the transfer of disparate data center physical infrastructure to the QRC virtual infrastructure platform. Each disparate data center will need to be taken one-at-a-time on an individual basis to ensure user service levels are maintained. ETC - March 2010 to 2012.

3. Decommission Disparate Data Centers - As services are transferred from the old physical to the new virtual infrastructure, the old physical DC will no longer be needed and will be decommissioned. ETC - March 2010 to 2012.

Sources of energy efficiency

This change in operating environment will create major efficiencies in the energy required to support Onsite Customer's data processing demands. The method for measuring these efficiencies is derived by correlating the compute output measured in compute units ("CU") provided by the physical IT assets to the energy required to operate it. In the case of Onsite Customer, the CU most appropriate for the types of services provide by the IT plant is the bit output of the CPU cores as measured by the company's IT asset management information system. Onsite Customer uses only 64-bit cores so the output capacity of servers will increase by the number of cores per physical server and clock speed (GHz rating of the core). Coincidentally, the more cores and GHz in the server, the higher the heat density will be. The three main sources of efficiency are:

1. Higher CPU/core utilization

2. More efficient heat conveyance

3. Transfer of compute & heat loads to a more efficient central plant.

Measurement of Energy Savings

The following describes the method of calculation and measurement of savings:

1. Key savings formula – the savings formula will be:

$$kWh_{Saved} = (kWh_b/CU_b - kWh_a/CU_a) \times CU_a - R$$

Where:

kWh_b = Total kilowatt hours consumed by data center QRC, pre-project. Loads will include IT Equipment (servers, routers, disk drives, etc.), Heat Transfer Equipment (CRACs, fans), power supplies transformers, lighting and process cooling (chilled water production). The kilowatt hours will be measured and/or calculated over a baseline period that reflects the efficiency of current operations.

CU_b = Total compute units produced for the baseline period. Compute units in this case will be measured as bit output from the CPU cores in the QRC data center. Bit output is a count of the actual bits produced by the CPU cores and represent the production of compute services for the user base. In Onsite Customer's case this metric correlates well with the workload demanded by the user base. It can be measured by Onsite Customer's IT asset management software, CA e-health.

kWh_a = Total kilowatt hours consumed by data center QRC including the container expansion. The metric will include the same loads as in kWh_b above. The kilowatt hours will be measured and/or calculated over a post-installation period that reflects the efficiency of the post installation operations.

CU_a = Total compute units produced for the post installation period. Compute units in this case will be measured as described in CU_b above.

R = Total load of disparate data centers still in operation at the end of the M&V period. This will be calculated by the nominal IT load in the remaining data centers multiplied the PUE of the QRC data center as derived from the baseline metrics measured above. The inclusion of this metric into the savings formula will provide a strong incentive to decommission inefficient/redundant data centers as rapidly as possible.

M&V Period

Full implementation of this project is estimated to be 24 months. Over that period, the savings formula will be calculated at various intervals during the implementation period to verify the incremental savings achieved during the implementation. A final calculation at the end of the period will determine the final savings achieved for the project.

Payout of the incentive proceeds would be provided as follows:

20% upon installation of the QRC data center expansion

60% of savings achieved based calculated savings at the measurement intervals

100% of savings achieved based on calculated savings at the end of the project.