ROBERT FRANCES GROUP IT AGENDA
Business Advisors to IT Executives

www.rfgonline.com February 16, 2007

Mainframe Computing and Power in the Data Center

RFG believes power issues in the data center will continue to get worse before they get better. While IT shops and businesses have been the beneficiary of Moore's Law for years, companies are recently feeling the burning power of its "dark underbelly." Power consumption has been doubling every 18 months, to the point where IT managers cannot fill a rack with blade servers in a data center, since it is not possible to cool them. The good news for data center managers is that mainframe computers directly attack the issues of power and cooling, and do so in a significant way. First, for comparable levels of computation, mainframe systems take up dramatically less space. This directly aids in eliminating the need to spend tremendous amounts of money (and time) building new data centers to come up with the needed space. Second, mainframe computers are proving to be dramatically more efficient in cooling and power consumption, in most areas that can be measured.

Business Imperatives:

- While IT managers might think that they can simply add additional cooling or power to the data center to handle the increased power loads, it ends up being difficult to accomplish this, even if companies can direct resources to address the problem. First, the densities associated with some of the high-power chips on the market today cannot be cooled by existing technology when fully populating a rack. Second, the power utility cannot guarantee that it can provide all the power needed by the new fully loaded data centers, even if the systems could be cooled (which they can not).
- The integrated nature of the mainframe environment reduces the operational complexity, which leads to reduced operational costs. RFG believes this factor will become dramatically more pronounced as companies move to implement service-oriented architectures. The mainframe complexity is reduced in three areas: the infrastructure architecture, facilities management, and development and testing.
- When calculating the value of a solution, IT executives must measure the cost of implementing a solution against the risks associated with not having a solution, as well as the business (and, to a lesser extent, technical) benefits of such a solution. RFG often finds that these considerations are evaluated subjectively. When IT executives can evaluate these various factors in as objective a manner as possible, they are in a better position to make realistic trade-offs, and present this analysis to business managers in terms they are more likely to understand and support. RFG finds the TCO analysis of operating mainframes, especially with respect to power and cooling costs, provides a significant operational advantage.

The Cost of Power

Even if it were possible to cool a rack of blade servers, the cost to power them is becoming prohibitive. One rack of energy-dense servers can consume up to 20 kilowatts of power. At today's energy prices, that can equate to $21,000 per year per rack just for the electricity. Conservatively adding another $10,000 per year for the electricity to cool these systems brings the annual bill to $30,000 of electrical costs per rack. The bad news is that these energy and cooling costs are only going to rise.

If the energy costs were not bad enough, the rise in computation and use of blade and rack systems for distributed computing is exacerbated by the increased use of data center space. At $20 or more per square foot in most of the areas that data centers are located, real estate is becoming a significant cost as well. RFG is seeing some large companies build data centers in low cost real estate areas for this reason alone.

The problem with distributed systems is that most IT planners are finding that they cannot use all of the data center space allocated for distributed computing. Most planners had worked with facilities managers

on the assumption a fully loaded rack would consume 3 to 5 kilowatts of power. The increased density of current servers has led to a situation where most racks can only be filled to at most 50 percent capacity. The direct impact on operations is that data centers effectively run out of space, due to cooling limitations, when they reach that 50 percent capacity point.

Mainframe Power and Cooling Advantages

The good news for data center managers is that mainframe computers directly attack the issues of power and cooling, and do so in a significant way. First, for comparable levels of computation, mainframe systems take up dramatically less space. This directly aids in eliminating the need to spend tremendous amounts of money (and time) building new data centers to come up with the needed space. Second, mainframe computers are proving to be dramatically more efficient in cooling and power consumption, on every front that can be measured.

Mainframe systems consume less power, both in absolute and relative terms. Typically, mainframe power densities are less than half of those of current rack and blade distributed systems. When looking at like workloads, the amount of energy consumed falls precipitously, in some cases the costs associated for power needed for an application are reduced by a factor of 600. Second, the power and cooling systems in mainframe architectures optimize cooling efficiency and power distribution, which significantly increase system uptime, while lowering the relative energy needed for cooling, further lowering operational costs.

Power and Cooling Issues Affect Staffing

RFG has found that increased failure rates associated with inefficient power and cooling decreases the full-time equivalent (FTE) ratios for system administrators. RFG has found that the failure of rack-based systems is significantly higher on the systems at the tops of racks. Cool air is typically delivered from the bottom of floor tiles and forced up the front of the rack via air pressure. The lower systems in the racks get chilled air, but the air available at the tops of racks is not able to cool the systems located there effectively. RFG has found this effect is aggravated when a rack has gaps in between systems. While many administrators create gaps in an effort to allow for better airflow, in fact these gaps create local current eddies which locally circulate air, actually preventing cool air from circulating up the rack. The higher failure rates associated with these systems detract from routine administrative tasks, which increase the overall number of administrators needed to attend to these systems.

Additionally, the complexities associated with distributed environment further work to reduce system administration ratios. RFG has seen an increasing use of virtualization technologies, such as EMC's VMWare to help consolidate more applications on single servers. While this can reduce the number of physical systems, RFG has found that this does not decrease the number of operating system images, and in fact adds the requirement for having an administrator that can configure the VMWare environment.

RFG has found that increased failure rates can also increase the stress on IT staff, which leads to reduced efficiency and decreased server to administrator ratios. In talking with IT executives, especially with those in mixed mainframe/distributed environments, the administrative ratios for mainframe systems administrative tend to stay constant. So, as the workload increases, there is not a concomitant need to increase the number of administrators on the mainframe. There is some increase in complexity as a result of managing mainframe logical partitions (LPARs), but the relative rate of management overhead increase is much lower than in distributed environments.

Complexity Associated with Distributed Environments

Additionally, the integrated nature of the mainframe environment reduces the operational complexity. RFG believes this factor will become dramatically more pronounced as companies move to implement service-oriented architectures. The mainframe complexity is reduced in three areas: the infrastructure architecture, facilities management, and development and testing.

Because the different application tiers can be represented on different LPARS on the same physical system, the complexity of modeling interconnections and anticipating application performance in production is simplified. As SOAs become more complex as they grow to address real business problems, the ability to accurate model these environments and guarantee the necessary level of QoS will be challenging. Mainframes have a distinct advantage in these environments, since all the tiers of the SOA architecture are in the same physical system, with low-latency interconnectivity between the elements.

These same advantages will accrue for development and testing, since any part of an application that is developed on an LPAR can easily be ported into the production environment. Applications that are developed in the distributed world have to undergo load testing, stress testing, and integration testing in separate environments, often with different staff. (Mainframe applications must undergo similar rigor, but sometimes in more consistent environments and with more consistent staff.)

A significant amount of time is spent porting the application to the different physical environments. Most of the time and the complexity associated with these migrations are eliminated in the mainframe environment. The fact the mainframe systems now employ dedicated processors for Linux and database environments add the capability to do this development and testing in the emerging open environments many companies are moving towards.

Case Study: Reducing Power and Space by Switching from Unix to z/OS

A large national retailer was having a problem with both the heat and space of its incumbent systems. The retailer was managing 200 Sun 4800 4-way servers. The systems were measured to consume 608 kilowatts and generated 2,074,000 BTUs of heat. Based on its geographic location and cost of electricity and cooling, the systems cost $30,000 per month to supply the electricity to run the systems and the air conditioning necessary to cool them.

Because of this large recurring cost, the retailer was looking at compute alternatives that could save on this $360,000 annual bill, which was expected to go up, as a result of skyrocketing energy costs. One of the alternatives they considered was putting their workloads on mainframe computers.

The retailer evaluated the IBM z9 54-way system. It was able to meet the computing requirements in the 200 Sun 4800 systems. This, considering that half the servers were idle as backup, and the remaining servers were averaging about 20 percent CPU utilization. While a single z9 54-way system drew 18,300 watts, compared with 3040 watts for one Sun 4800, the single box was able to do the work of the 200 systems. Therefore, the comparable performance metrics for real workloads ending up being 608,000 watts for the total Sun systems, compared with 18,300 watts for the z9, or 3 percent of the comparable power draw. The BTU comparison was 62,220 BTUs for z9, compared with 2,074,000 BTUs for the Sun systems. The difference as it related to the monthly power bill at $.05 per KWH came to $905 vs. $30,165, for an annual electricity savings of $351,120 or a reduction of 97 percent.

In addition to the significant electricity cost reduction, the retailer also found a significant space savings. The total space of the z9 system took one-tenth the floor space. RFG has found similar metrics when comparing mainframe space and power utilization to other distributed systems.

Sun/IBM Server Energy Comparison at National Retailer				
Server	**Total Watts**	**BTUs**	**Monthly Cost**	**Annual Energy Cost**
Sun 4800 4-way, 200	608,000	2,074,000	$30,165	$361,980
IBM z9 54-way 1	18,300	62,220	$905	$10,860
% Difference	(97%)	(97%)	(97%)	(97%)

Source: Robert Frances Group (RFG)

RFG believes mainframe computing platforms have many of the characteristics that will ameliorate, if not eliminate, the current challenges data center managers face with power and cooling. First, mainframe power consumption and heat characteristics are, for many companies, the most efficient servers in the data center. This is true in an absolute sense, where the energy per square foot is lower than any data center system measured by our clients. More significantly, this is massively true in a relative sense, when comparing power used per transaction. On a total workload throughput basis, mainframe system power consumption is almost negligible when compared with distributed systems on a power per transaction basis. As power and cooling costs continue to rise, IT executives should reevaluate mainframe computers total cost and overall value in reducing data center operations costs. See the RFG Research Note "Data Center Transformation: Observations and Recommendations" and the RFG Research Briefs "Vendor Response to the Power Problem in the Data Center" and "Data Center Cooling Best Practices.")

RFG analyst Jerald Murphy wrote this Research Note. Interested readers should contact Client Services to arrange further discussion or an interview with Mr. Murphy.