ROBERT FRANCES GROUP
Business Advisors to IT Executives
120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Powerful Incentives:

Using IBM System z to Realize Significant Operational Cost Savings

Jerald Murphy
Senior Vice President
Director of Research
Robert Frances Group

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Table of Contents

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Executive Overview

One of the recent phenomena to hit (or perhaps collide) with the data center is the impact of power and cooling on operational costs and constraints. 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.

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 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/ 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+ per square foot in most of the areas that data centers are located, real estate is becoming a significant cost as well. 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% capacity. The direct impact on operations is that data centers effectively run out of space, due to cooling limitations, when they are only at 50% capacity.

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.

First, 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.

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

On top of the directly measurable power and cooling benefits, second-order effects related to mainframe computing are equally significant. First, significantly lower failure rates associated with mainframe computers relates to lower management costs. Second, the inherently consolidated mainframe environments reduces development and deployment costs, since the time and complexity associated with testing and deploying applications from the development to production environment is greatly reduced. Third, the security associated with mainframe environments is historically much higher than those in distributed systems.

Put together, lower energy costs, better space utilization, reduced management costs, simpler deployment, all unified with centralized security combine to make mainframe environments significantly more cost effective than distributed processing alternatives.

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Business and Technology Scenario

It is no secret that CIOs are under constant pressure to reduce operational costs. Personnel costs continue to rise, the cost of real estate rises, material costs rise: all costs continue to rise. Most significantly, the costs of power and cooling in the data center are rising.

Moore's Law has been very good for IT managers, so far. For the last twenty years, processor speeds have doubled, while lithographies have halved about every eighteen months. This has make computers dramatically more productive over time, for essentially no incremental cost to IT. Of course, overall costs have risen, as companies use computers for more tasks. The cost that has been largely been ignored until recently is the energy costs related to operating these computers, and the closely related costs associated with cooling this equipment.

The Power and Cooling Conundrum

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 not so simple to accomplish this, even if companies have the money to do this. 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 they can provide all the power needed by the new fully-loaded data centers, even if the systems could be cooled (which they can't).

This issue prevents many companies with a difficult dilemma.

Space vs. Power and Cooling: Data Center Scylla and Charibdys

The problem most companies are faced with today is whether and how they can afford to add new compute platforms in their data centers. On the one hand, additional systems added need additional power, which may not be available. On the other, they made need more space for the compute systems, which itself may not be available. Additionally, most companies do not have adequate chargeback mechanisms in place to accurately handle IT costs associated with power and cooling, since these costs have typically been borne by the facilities management group.

Space Issues

Real estate is expensive; under utilized space creates tremendous inefficiencies RFG has had several clients say that the cost of adding an additional rack server to the data center is $10-20 million. While there is physical space left in the data center, it cannot be used since the space left doesn't have the capability to be cooled. Because of this, the space that was planned to last for 7-10 years in capital planning is gone in 4-5 years. This means that another data center needs to be built to handle incremental computing requirements, which typically runs from $10-20 million, depending on the geographic location and size. IT executives that are able to maximize the use of existing space will save these capital dollars.

Power and Cooling Issues

The problem with using all the existing space for computation is this: if a rack is filled with current rack or blade server technology, it can’t be cooled. Most existing data centers managers that RFG interviewed state that their cooling capabilities are currently limited to 10kw per rack. Most racks have been designed with typical power densities of 3-5kw. When systems are populated above this density, elements at the top of the rack have low MTBF.

Power and cooling companies have developed technology that will allow racks to achieve cooling densities of 15-20kw per rack. Companies such as Liebert and APC make higher density cooling, but these units occupy a rack next to the one to be cooled, so net space available isn’t any better than if a rack was half full. The net result of this is that real estate that would have been used for servers is now being used for cooling, so the net space available in the data center for computation has not changed with this technique.

Even if these systems could be cooled, there often isn’t power available from the utility company to handle these demands. RFG has spoken with several large financial services companies who have decided they needed to be responsible for primary power generation, since the utility company could not guarantee peak delivery. Not only is there the additional costs associated with building power generation systems, but additional administrative costs associated with site surveys, zoning approval, and environmental oversight that have to be addressed in order to build these large power generation facilities.

With all these power, cooling, and space issues that seem to directly work against one another, the future situation for the data center appears to be on track for inexorable geometric growth in costs related to energy, cooling, and space.

Given this situation, RFG has found that the use of mainframe computers in the data center directly attacks each of these problems, yielding dramatic improvements in both operational and capital costs.

How Mainframe Computers Solve These Powerful Issues

RFG has found that mainframe computers have ideal characteristics for dealing with the triple-threat complexities of power, cooling, and space in the data center. In addition to having features that directly address these challenges, the tertiary issues associated with improving these areas have a multiplicative effective on overall data center efficiency.

120 Post Road West, Suite 201, Westport, CT 06880   Phone: 203-429-8950   Fax: 203-429-8930

Minimum Power Drain

Mainframe systems have been designed to maximize overall energy efficiency in several areas. First, the power draw of the processors themselves has been designed for lower power utilization. Second, the power frame itself uses energy very efficiently. Mainframes are designed with a central AC/DC power converter, which operates at over 90% efficiency, compared with many existing rack server power converters which operate at 70%. As a practical example, in one large insurance company RFG interviewed, the difference this equated to in energy consumption was over $300,000 for energy dedicated to distributed systems, not including cooling, compared to less than $150,000 annually for power and cooling dedicated to mainframe systems. The total transactional throughput of the mainframe systems was greater, at less than half the energy and cooling costs.

UNIX and x86 power Issues

While IT executives that RFG has spoken with like the advantage of open systems flexibility and adaptability, and some companies like the lower initial capital outlay normally associated with Windows and x86-based platforms, the power issue with these systems is becoming a major concern. Most of the CIOs of large financial institutions RFG has talked with have stated that the power and cooling issues associated with these systems has become their largest concern. Also, while the manufacturers of these chips and systems are planning on building more efficient next-generation systems, many of these are not yet available.

Maximizing Space Utilization

RFG has found that mainframe systems take up much less space than comparable rack-based or blade systems. In a typical data center of RFG clients, the space occupied by mainframes constitutes less than 10% of the total data center space. When considering the space occupied based on computational throughput, the relative value of mainframes per square foot becomes significant.

Most of the clients RFG interviews have a tremendous amount of spare capacity available on the mainframe systems in house. One client RFG interviewed was using 1000 MIPS of compute power on the mainframe, and the system they had in place was capable of delivering up to 6000 MIPS. This means that when the company needs to add incremental compute capacity, already existing engines can be turned on, with very little incremental power expenditure and no additional floor space needed. For this reason, the mainframe has been an excellent system for them in order to allow growth on demand for their business with no additional facilities cost.

While the direct comparison of power, cooling, and space is fairly obvious and easy to measure, additional benefits related to these issues can be seen as well.

Reduction in Personnel Costs

Systems Administration

RFG has found that increased failure rates associated with inefficient power and cooling decreases the 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, where the air available at the tops of racks is at a lower temperature and is not able to effectively cool these systems. RFG has found this effect is aggravated when a rack has gas 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 environment

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

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

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. 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.

Other factors

In addition to the elements mentioned that directly relate to mainframe advantages, other benefits that are also seen in mainframe environments include:

- Power Conversion
  Mainframe systems typically have over 90% efficiency in converting AC power to DC power. This compares with existing distributed systems that are usually between 70-80% efficient.
- Code Conversion Costs
  By adding new systems to the mainframe, it eliminates the need to spend time and money converting legacy mainframe code (COBOL, IMS, etc.) onto distributed systems. This is expensive and risky, since most of this legacy code is for mission-critical, revenue producing business applications
- Security
  Mainframes have a very strong reputation for security, with few virus problems, encryption, strong authentication, built-in PKI, etc.
- Efficient Resource Utilization
  Most of the distributed systems of companies RFG interviewed have dedicated servers with applications that run between 10-20% CPU utilization. This is one reason for much of the server consolidation. Mainframe systems typically run over 80% utilization, so maximum use it obtained for what has been purchased.

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Now that the benefits of mainframe computers compared to distributed blade and rack systems has been described, a cost benefit analysis will be reviewed to show how these issues financially impact companies that RFG has interviewed.

The ROI of Mainframe Power and Cooling

Value calculation

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 tradeoffs, and present this analysis to business managers in terms they are more likely to understand and support.

Consideration Factors

Total power

RFG got information from clients on the total energy used by mainframe and distributed applications that were running in their data centers.

System Space

Clients were asked about the amount of space taken up in square feet by mainframes, rack systems, and the total data center space.

Cooling

Cooling costs were estimated by data center managers, since those who were interviewed were not given separate cost allocations by facilities managers for cooling. The estimate was generally agreed to be 50% of the energy used by the computing systems was needed for the air conditioning systems.

Power per application

Power per application was calculated on the mainframe using the MIPS dedicated to an application as a fraction of the total MIPS on the system in production. Application power for distributed systems was determined on systems with dedicated servers for an application.

ROBERT FRANCES GROUP
Business Advisors to IT Executives
120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Figure 1 Sample IBM System z Value calculation

Mainframe power analysis				**Distributed System power analysis**	
	Z App 1		Z App2	2	Processors / server
Total power for mainframe	10500	Watts		233	Watts / processor
MIPS for application	0.20	MIPS	6		includes power
Total MIPS in mainframe	2600		2600		supply, fan, etc.
Power allocated per MIPS	4.04	Watts/MIP	4.04		
	0.81	Watts/app	24.23	466	Watts per application
Cost per kilowatt hour	$ 0.10		$ 0.10	$ 0.10	
Operational hours per year	8760		8760	8760	
Cost per year for app energy	$ 0.71		$ 21.23	$408.22	
Cooling Costs	$ 0.35		$ 10.61	$204.11	
Total yearly app costs	**$ 1.06**		**$ 31.84**	**$612.32**	

Figure 2 Average energy consumption by system

Gross average energy costs per system	Power Costs		Percent
UNIX Version A	$165,000		29.7%
UNIX Version B	$75,000		13.5%
UNIX Version C	$110,000		19.8%
Linux	$32,000		5.8%
Windows	$52,000		9.4%
z/OS	**$122,000**		**21.9%**
Total Energy Cost	**$556,000**		

Taking into consideration the various factors that drive power consumption in the data center, mainframe systems are clearly superior in total energy utilization. This is both in terms of total energy used on a gross basis, but more importantly, on a cost per application basis. While determining application comparisons between distributed systems and mainframes is not exact, even by the most conservative estimates mainframe systems have dramatic energy and space efficiency benefits that cannot be ignored. RFG recommends that IT executives take a close look when comparing mainframe systems to distributed systems at total transaction throughput and calculating the energy per transaction to get a better estimate of the relative costs between mainframe and distributed applications for a true sense of energy efficiencies associated with each of these systems.

ROBERT FRANCES GROUP
Business Advisors to IT Executives
120 Post Road West, Suite 201, Westport, CT 06880   Phone: 203-429-8950   Fax: 203-429-8930

Case Study I

Reducing power and space switching from UNIX to z/OS

A large national retailer was having a problem with both the heat and space of its existing systems. The retailer had 200 Sun 4800 4-Way servers that it was managing. 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% 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% 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%.

In addition to the significant electricity cost reduction, the retailer also found a significant space savings. The total space of the z9 system took ten times less floor space. These calculations didn't even include the cost of power, cooling, and space associated with the external disk needed to support the Sun systems.

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%)

Figure 3 Sun/IBM Server energy comparison, national retailer   Source: Robert Frances Group

Case Study II

Nationwide Insurance Adopts Mainframe virtualization to save

Nationwide Insurance was leveraging Linux for their Web application serving. They looked at the rising costs of dedicated servers to each Linux instance and decided to extend their existing mainframe operation to include those Linux servers as z/VM virtual guest images instead. Their introduction preceded Nationwide's rebranding via a new Super Bowl ad in January 2006. Given one month prior to the Super Bowl to set up a Web presence to include the new ad, Nationwide was able to test 22 times greater than their actual needs, prior to going live. During this two week period, they were able to rent the additional capacity, without installing any new machines. They dynamically turned on and off capacity within their z9 mainframe.

Virtualization provided a number of significant savings to Nationwide. There was a 50% reduction in virtual server images as they were able to run the mainframe at 70% average CPU utilization vs. around 20% when they had dedicated servers to each Linux image. The mainframe used 80% less data center floor space than their distributed operation. This resulted in significant energy savings, similar to those described in the retail customer case study. There is a significant middleware cost savings due to the reduction in CPU's necessary to meet their computing needs. Availability has been improved due to server redundancy, but there was also a 50% reduction in hardware and operating system support costs due to improved reliability of the mainframe. Each dedicated Linux server may have had a mean time between failures of 20 years, compared to greater than 40 years for the mainframe. With so many individual servers, a great amount of time was spent repairing the individual server images in the distributed environment.

Nationwide found that the capabilities of zSeries plus z/VM provided better qualities of services and application isolation compared to other Linux virtualization techniques. Their largest transition was convincing end users that a virtual server image was as good as, if not better than, a stand alone server image. They found that each of their cost structures improved with mainframe virtualization: data center energy and real estate, server capital costs, server support personnel costs, software license costs, software management personnel costs, software maintenance costs, application support costs and disaster recover support.

120 Post Road West, Suite 201, Westport, CT 06880 Phone: 203-429-8950 Fax: 203-429-8930

Conclusion

The costs of data center power and cooling have grown exponentially in the last several years. Currently, power and cooling costs consume from 30-40 percent of most operational budgets. Left unchecked, the costs related to power and cooling will be more than the costs of purchasing the equipment being run. With the cost of energy continuing to rise, this threshold is likely to be breached very soon. In large energy cost locations, such as the Asia Pacific region, this threshold may have already been reached.

To date, IT executives have been reacting to this situation with point solutions that provide only temporarily relief. Consolidating applications on one system helps, as do more efficient cooling and better power management practices. However, like adding sandbags to a broken levy, it is only a matter of time before IT and facilities managers are in a situation where they cannot cool the systems they have, and must make difficult decisions on investing in new facilities, or restricting the growth of business. Neither choice is good, since building new facilities takes more time than the business has, and stalling corporate growth is a recipe for business failure.

Mainframe computing platforms have most of the characteristics that will ameliorate, if not eliminate, these challenges. First, mainframe power consumption and heat characteristics are the most efficient elements 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.

In addition to the clear power and cooling savings, the additional benefits that accrue to mainframe usage complement the operational savings from power and cooling. Application development and implementation costs are reduced with the mainframe's ability to easily move LPARs created for development and test to be moved into production with no infrastructure configuration changes needed. Administrative costs for managing a mainframe environment are dramatically lower as well, with fewer FTEs required for administration compared with distributed environments.

With the money saved from more efficient power utilization, investments can be made for additional capacity, training, and development; money invested that can be tied directly to company innovation and revenues.

Mainframe systems are clearly superior in handling the growing power and cooling issues in the data center. Taken together with the cascading efficiencies and cost savings that are a direct result of these attributes, mainframe systems stand out in the unique ability to realize true business value from data centers.