



Summary Insights: Unleash the Power of OpEx To Transform Data Centers and Kill CapEx

RFG Perspective: The two main drivers of data center transformation today are AI (primarily generative AI) and sustainability. Both shifts will have substantial effects on hardware acquisition strategy and data center design including requiring evolutionary operational metrics and added funding. GenAI capabilities will add huge costs across the categories of compute, storage, network, and power, necessitating IT executives understand and appropriately allocate limited resources to achieve business objectives. Moreover, existing CapEx funding appropriations for IT genAI hardware acquisitions is highly unlikely to cover these needed allocations, which will require line-of-business and enterprise support contributions on an ongoing basis. Similarly, ESG directives may also require some degree of CapEx funding; however, sustainability initiatives can prove self-funding if the acquisition and operational metrics and processes are amended to holistically approach the challenge via an OpEx model.

INTRODUCTION

Most IT data center executives use a CapEx model for their method of hardware acquisition as it requires less effort and less personal risk for IT executives. But it is the least desirable means of funding data center transformations. CapEx falls short for two reasons: the CapEx model has IT competing with other parts of the business for scarce capital funding which exposes it to the risk of funding shortfalls, and it reduces the flexibility to change as the project or initiative changes. One advantage of the CapEx model is that IT executives do not require a good asset management system in place (a common data center shortcoming and a, if not THE, leading reason CapEx is still preferred for corporate-owned equipment), which is a requirement for the OpEx models if one hopes to keep costs under control.

On the other hand, while the OpEx models require a better understanding of financial analysis and cost structures, it allows for greater flexibility and reduces operational risk. There are multiple OpEx models to choose available including hardware leasing, hardware and services rental models, consumption-based pricing, and more novel approaches that include a variety of technology and related services within a single price point. Cloud service providers (CSPs), colocation providers, and hardware suppliers offer multiple OpEx versions, each has its pros and cons.

The Panel

The panelists on the call were:

- Tony Grayson – GM, Compass Data Centers
- Pat Bodin – Global AI Architect, Cisco Systems



- Deal Daly - Advisor, ATEM Group
- Adam Braunstein, Principal and Director of Platform Computing Strategies, RFG

Rack Density and Other Facilities Concerns

IT lifecycles are fractionally shorter than the facilities lifecycles. Facilities teams plan and build on the assumption that the data center will be operational for ten to twenty years or more while IT works in one-, three-, and five-year terms. This disconnect has long-term implications as radical changes in technology capabilities, density, workload types, and other factors may not be easily incorporable into Facilities teams' initial designs. Most data center rack, aisle, power, and HVAC buildouts are Facilities' planning responsibilities and attempts are made to design for the long-term – at least in theory, as that often fails to come to fruition. As a result, data centers and hyperscaler facilities have different architectural requirements and design points. Most data center racks have rack power densities of less than 15 kVA – with six to eight kVA being the current standard – whereas hyperscalers have built their facilities with densities of 40 to 50 kVA moving to 80 and 100 kVA racks. The greater the rack density, the less space potentially required for containing the required equipment (non-power factors notwithstanding). Conversely, the greater the density, the greater the blast radius and the higher probability for catastrophic failures resulting in business interruptions.

IT executives cannot easily switch from low-density racks to high-density racks without agreement from Facilities and the supporting work effort required. Facilities may not be able to accommodate the upgrade to the power as the energy utility may not be able to deliver the desired increase in the power supply. IT executives need to understand the long-term constraints going out five years or more. On the other hand, the hyperscalers have gone from designing campuses of 40 megawatts to upwards of 500 mW. The cloud providers and other hyperscalers mostly use an OpEx model for acquisition and depreciation purposes. It is far easier to fund these large campus facilities and hardware investments on an OpEx basis than it is to fund it through a CapEx model. Furthermore, this approach enables hyperscalers to grow more quickly, as the OpEx model requires less upfront capital.

Leasing and the OpEx Models

Colocation providers operate using both retail and wholesale models. In the wholesale model, which they use when providing space for hyperscalers and some large enterprises, pricing is by kilowatt. The retail model is used with data center executives who want to pay based on square footage required and are density focused. Retail buyers will end up paying for space and power. Most data center executives lease the colocation space but still own the equipment that fills the space. By utilizing a colocation provider, IT gets flexibility for growth but by purchasing the IT equipment, IT constrains its growth opportunities. IT executives should switch to a matching OpEx model that will improve both technology performance and compute capability per square



foot on an ongoing basis. This OpEx model alignment across space and technology will be easier to fund and budget on an ongoing basis while maximizing investments throughout.

Enterprises of all types, including Financial Services firms, employ leasing strategies to reduce spend rates and encourage regular refresh. The biggest blocker, by far, to leasing and other technology OpEx adoption is the lack of effective asset management (deployment, location, migration, retirement, end-of-life, etc.). IT executives lacking skill sets and/or FTE support to address these challenges will need to address shortcomings either internally or via one or more third parties. IT executives should not shy away from the OpEx model -- even the government loves leasing and using colocation sites. One of the advantages of using a colocation facility is that it tends to be less expensive than using a corporate-owned data center. This is because colocation firms have excellent credit ratings and, in many cases, can obtain better credit terms than corporations or government agencies. Lower cost of capital means less risk. The effective use of leasing along with asset management can result in lower on-premises and/or colocation site costs than through cloud provider services for large enterprises.

Consumption-based pricing is another OpEx model being used by several vendors that can be advantageous to IT executives. The contracts are self-optimizing, which relieves the burden of doing frequent capacity planning exercises to get best pricing from the IT team. This method may prove to be a little more expensive per unit cost than a fixed price contract. The value is realized in that it does not require a long-term price commitment -- which could be based on erroneous planning assumptions. Vendors, such as HPE and Lenovo, will provide a full data center service capabilities for contracted OEM (and some partner) systems: consumption-pricing plus capacity planning with provision of additional hardware when required.

Regardless of the OpEx model, if IT is currently using a CapEx model then IT executives must work with the CFO to agree to the change. The issue here is often less explanation, justification, and faith in the financial model and more related to trusting that the IT team can transition to the new model without business disruption and/or radical cost overruns. The challenge that CFOs experience with IT executives is related to the frequent lack of adequate asset management and supporting services, as mentioned before, which results in paying excessive OpEx fees to the vendors that thereby negate OpEx advantages.

AI

The initial CapEx investment in AI is typically significant and is accompanied by a development cycle likely exceeding one year. It is therefore essential to plan early and thoroughly while wholeheartedly acknowledging that AI deployment is neither a quick-fix solution nor without numerous roadblocks and obstacles to overcome given its newness and the lack of institutional knowledge. This will be a strategic, long-term investment with a lengthy ROI and associated risk. IT executives must also understand that the complexity of the technology means that it



requires significant computing power (CPUs and GPUs) whether on-premises or off-premises, extensive data sets for training the AI, and expert knowledge to maintain and adjust the system.

Because AI consumes large amounts of computing power, air cooling vs. liquid cooling comes up frequently in these discussions. Hyperscalers and colocation providers, for the most part, are still air cooled but have adopted liquid cooling in certain instances. Most of the air cooling is for sidecar augmentation. Liquid sidecars have a dielectric fluid that goes to a heat exchanger, which is in turn cooled by a direct exchange with the facility's HVAC system. These sidecar systems are currently being deployed by the big cloud providers for their large language model (LLM) deployments.

To avoid this major CapEx expense, IT executives should move to a phased OpEx approach to delivering AI, especially genAI. In this way, one can pay for "just-in-time" system costs and usage rather than a large upfront lump sum payment. As additional equipment, space, or cloud services are needed, they are acquired thru additional OpEx acquisitions. CSPs are in the process of making the procurement of AI services easier by creating AI as a Service options.

Corporate Alignment

One of the major challenges for many IT executives is that their world view is different than that of business executives. IT executives think of their operations as cost centers and focus on the operational goals – primarily performance and cost. That is what they are measured on annually; and, therefore, it is what they focus on. Business executives think differently and are measured on revenues, profits, and customer retention and growth. IT executives tend to think mostly about the current year and not future years, while line of business executives need to have three- to five-year plans in place. And, unfortunately, most IT executives do not bother to understand the long-term business plans and work collaboratively with their business peers to support these needs. This needs to change if there is to be an effective alignment of resources.

SUMMARY

Data center transformations can be expensive, multi-year projects. Corporate and line of business executives may agree with the vision but extracting the funds tends to be the rub. IT executives need to sell their internal customers on the value of the transformation and how it can be obtained without breaking the budget.



RFG POV: AI, especially genAI, solutions are being pushed by Boards, corporate, and line of business executives as the new answer to today's and tomorrow's business problems. About half believe it will drive added revenues while the other half believe it will improve productivity or reduce fraud and thereby reduce costs. ESG mandates are also pressing these executives to implement various sustainability solutions across all sectors of the organization – with IT beginning to find it is no exception.

IT executives cannot take a defensive position and must find a way to meet current and future business demands within budgetary requirements. For many IT executives, this will require a change in their business models and how they interface and collaborate with the corporate and line of business executives. IT executives should work collaboratively with their peers to develop one-, three-, and five-year operational plans on how they can transform their data centers to meet the needs of the business – both in terms of outcomes and expenses.

Additional relevant research and consulting services are available. Interested readers should contact Client Services to arrange further discussion or interview with Cal Braunstein, CEO and Executive Director of Research.