



Using the Cloud to Maximize Return on Investment

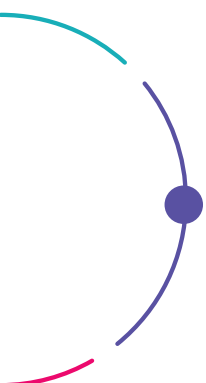


Table of Contents

Overview of IT Expenses	3
Capital Expense Versus Operating Expense	3
Cloud Cost Savings	6
The Hidden Cost of Data Inefficiency	6
Storage Optimization	8
Automated Backup	9
Developer Efficiency	9
Ease of Data Management	10
Summary	11
Learn More About Panoply.io	11





Overview of IT Expenses

In a traditional IT environment, corporations spent 70% of their resources just to “keep the lights on”, as the saying goes. Those types of operations require a long list of expenses including - pure hardware lifecycle issues like replacing servers every three years, adding storage (you’ll learn more about the cost of storage later), network equipment, and data center power and cooling. For an organization that wants to protect its data, this means multiplying these costs by two or three times to provide redundancy across multiple data centers. In many cases, it also means having staff in each of those locations to support that hardware.

There’s a significant financial outlay before even considering expensive support contracts on both hardware and software—storage arrays require expensive support contracts, as does relational database management software. Many large organizations pay tens of millions of dollars in support costs annually.

Cloud computing changes some of these paradigms—in most cases, you can “rent” hardware and software, transferring what was traditionally a capital expense to a more palatable operating expense. In the next section, you will learn about how that can be beneficial to your business.

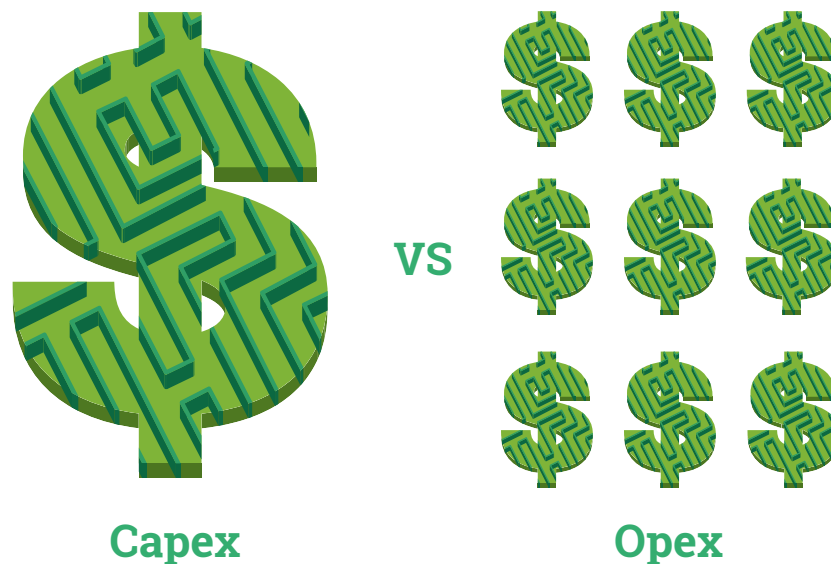
CAPITAL EXPENSE VERSUS OPERATING EXPENSE

A capital expense, or CAPEX, is an accounting term that refers to physical assets that the business amortizes over several years. This means that, from a tax perspective, your business cannot deduct the full cost of these assets in the fiscal year they are acquired. For example, if you build a data center and fill it with storage, servers, and network gear, you can deduct only a percentage of that cost each year—this is called amortization. This is because your company is expected to derive value from the asset over several years.

CapEx spending means different things to different companies—in the case of small start-ups, it is just not practical to have the amount of capital on hand to make such a major investment. Before cloud computing, it was tough to stand up the required amount of infrastructure to host their solutions. For larger public companies, increasing capital expense means impacting budgets, return on investment, and

potentially stock price. These are all things that are of primary importance to the CFOs of both companies.

The other side of organizational expenses is operational expenses, or OpEx. These are a little bit easier to understand and qualify. Unlike CapEx, which a business amortizes over multiple years, operational businesses expense hit the books in the year of the cost and, for taxable entities, create savings in the form of expense deductions. Your company derives the value from the expense immediately. Some simple examples of this are travel expenses, or as we are talking about here, cloud computing services.



There is no hard-and-fast rule about what should be OpEx versus what should be CapEx. However, if you think about options involving the two, you can form a better understanding. For a smaller firm, the notion of purchasing all computing services as OpEx provides several benefits. It drops the barrier to entry and democratizes a lot of services that were previously available only to much larger companies. For example, in the past, the concept of a massively parallel processing engine like Panoply.io required the purchase of a dedicated appliance from a vendor like Oracle, Teradata, or Microsoft. At a minimum, this required an investment of hundreds of thousands of dollars of hardware, with a multiplier for consulting fees for implementation, and

software costs on top of that. Above that is the added cost of annual support on both the hardware and software, which on a system like that alone can add up to hundreds of thousands of dollars.

Another example of technology that was democratized by cloud computing is the ability to have a second data center for disaster recovery (DR). While some organizations can tolerate some degree of downtime without impacting their business, in our increasingly connected world, many applications need to be able to survive the loss of a primary data center while mostly maintaining uptime. Before cloud computing, this meant either renting expensive co-location space or building and managing a second data center in a second geography. Now a company can inexpensively provide DR using cloud resources and scale them as needed in the event of a disaster.

Another way to look at this CapEx and OpEx discussion is to think about a department within a large company. The department may wish to do a data analytics project; however, the associated CapEx costs will require executive approval and may need to pass through several budget cycles. Additionally, as mentioned earlier, the lead time associated with a major IT project could double the time it takes to get the project off the ground. With a cloud solution, if the smaller amount of funding required for the pay-as-you-go solution is in the departmental budget, the department can fund the project internally with a much shorter approval cycle. Once the connectivity to the cloud provider comes online, and an administrator makes the initial data source connections, the department can start to receive value from the project in very short order.



Cloud Cost Savings

The other benefit of moving workloads to the cloud is reducing the IT staff needed to manage systems or reallocating those resources to higher value tasks. A smaller organization might be able to get by with having only a couple of IT resources to work with the cloud provider and manage the connectivity to the cloud and the resources consumed in the cloud. In a larger organization, IT staff can be moved from jobs where they are merely responsible for keeping servers up and running and into higher value tasks like data management and helping business units derive value from their use of IT systems.

A larger organization would need to move the majority of its resources into the cloud to capture the benefits mentioned above. Moving to the cloud is a non-trivial strategy transition that may take many years to accomplish. However, for a smaller company, this shift is revolutionary and game changing. For a small and fast-growing company, the ability to have all of its human resources focused on mission-centric work rather than maintenance activity such as managing the upkeep of systems can be a huge benefit.

THE HIDDEN COST OF DATA INEFFICIENCY

In a data-driven world, one of a company's biggest assets is its data. The ability to gather metrics on customers, supply chain, and marketing campaigns provides the ability to make better decisions throughout the company, and support more informed strategic decisions. In larger firms, the silos created between departments may pose a challenge to bringing all of this data together in one place to perform an analysis. In a smaller company, the source systems may be a mix of software-as-a-service (SaaS) products and traditional software that the firm may lack the technical resources to bring together.

There are very real and costly ramifications of this sort of inefficiency. At some organizations, limitations such as those just described keep them from reporting their monthly sales until 30 days after month end. In that case, the organization would be unable to report its financial data promptly to companies that were interested in

them for acquisition, for example. Other companies may have useful data in place but may take days or months to get it into their target systems, which can impact their agility and flexibility in decision-making. The sooner you have higher quality information, the faster you can make critical business decisions.

The other challenge that organizations must overcome is the cumbersome and fragile nature of the ETL process. Every change to some downstream processes also potentially impacts other processes all the way back up to the top of the ETL process. This cascade means reconfiguring and redeploying code in addition to going through new testing cycles. While some attempts have been made to better automate ETL processes, they are limited in their adoption. Many organizations still rely on manual techniques or primitive scripts that can be greatly impacted by changes.



STORAGE OPTIMIZATION

One of the largest costs in any large IT organizations is storage. This is confusing to many people who aren't deeply involved in IT infrastructure and may go into a retail electronics store and see a 4-terabyte hard drive for less than \$200. However, enterprise-class storage, the kind that supports analytics systems, comes with a hefty price tag. In some cases, it can be up to \$3000 per terabyte. Why the dramatic difference in pricing? There are a number of reasons. For one, enterprise storage needs several layers of redundancy for protection in the event of hard drive or disk controller failures. Additionally, the storage in most larger organizations is network connected and requires the use of dedicated switches and expensive fiber optic cable to connect it to the servers it supports. These specialized storage devices require dedicated staff to maintain and support the storage array.

The presence of cloud computing changes the paradigm of storage. While in some areas of cloud computing (particularly IaaS virtual machines) the cost savings of using the cloud are debatable, storage is not one of them. Storage is markedly cheaper in the cloud, due to both economies of scale and engineering improvements from the cloud providers.

Cloud providers can offer impressive density at a low cost using a storage technique known as object-based storage. They reap additional efficiency because they are writing their own management software which is devoid of all the support and extra costs of legacy enterprise storage vendors. The object-based storage model also offers flexibility and allows cloud providers to simplify deployment of storage and build their own redundancy model without having to rely on a third party.

Object Storage vs. File and Block

File and block storage are well understood by most IT practitioners, especially those who deal with data. However, object storage is a new subject for many. The primary differentiator is that object storage neatly bundles up the data and associated metadata and stores it with a unique identifier. As opposed to traditional enterprise storage which often consumes proprietary hardware and uses RAID to protect data, object stores.

AUTOMATED BACKUP

Backups have come a long way since the days of the LTO tape being shipped to an off-site provider. Even though technologies have changed, the costs associated with storing backups have not changed that much. Backups are just additional data that needs to be stored, powered, and cooled. Administrators are also needed to ensure the success of backup jobs and perform storage capacity management.

One of the benefits of cloud-based storage is its nearly infinite capacity. A significant upside to using a platform as a service (PaaS) solution like Panoply.io that includes automated backups is the elimination of the need for the administrator to worry about backups, or the need for a third-party provider to provide off-site redundancy for the backup media. The redundancy is built into the cloud platform, and Panoply.io ensures that the data is backed up at regular intervals. There is no volume to run out of space; there's no tape library to fail—the backups just happen and are always there. Solid backups help business leaders sleep better at night while also reducing overall infrastructure costs.

DEVELOPER EFFICIENCY

Given the very manual and waterfall-based nature of ETL processing, it accounts for the largest portion of most traditional data analytics projects. Gathering, cleaning, and transforming data can take up to 80% of developer effort on a project. Unfortunately, this part of the project isn't even something that adds any business value; it is merely a part of the work required to start building any analytics system.

Imagine if, from Day One of your analytics project, your developers were focused only on building the best analytics queries and algorithms, and they had free time to focus on determining how to answer the next questions your business will have rather than merely cleaning and shredding data and worrying about a character change breaking that day's data load. With the automated data transformation that Panoply.io offers, that dream becomes a reality.

EASE OF DATA MANAGEMENT

A final area where Panoply.io lowers your TCO is the ease of access to your data. Once your data is loaded into your warehouse, it is time to use it. Panoply exposes a standard SQL endpoint that supports JDBC (Java database connectivity) and standardized ANSI-SQL. Your users and analysts can plug in any tool they want—Tableau, Spark, R, or any other standard data analysis tool. Panoply.io also delivers a set of extensible cloud-based analysis tools, which can lower your costs even more by allowing the consolidation of tools. This cloud analysis layer is completely open-source.

Other features included are the data management layer, which is a metadata editing tool for your data. Using this included tool from Panoply.io, you can identify your most popular queries and tune them based on your knowledge of the data. You can view and modify tables and data relationships such as primary and foreign keys. Additionally, if you want more control of your data and tuning, you can turn off Panoply.io's automatic data jobs or adjust them to align their timing to your business needs better.





Summary

Moving into a PaaS solution from a traditional on-premises solution can certainly be a leap of faith, as it requires giving up some measure of control. However, a powerful and robust platform like Panoply.io can provide higher velocity to insight, allowing your business to make better-informed decisions and improve growth opportunities. Additionally, your IT staff can be more productive because they will have a quicker time to results and less manual effort getting data into source systems.

LEARN MORE ABOUT PANOPLY.IO

Panoply.io is a cloud-based data management platform for analytics that streamlines time and value for data engineers, scientists and analysts – automating the full data stack without the overhead of preparing and modeling data, or managing infrastructure – cutting down development time by 80%. To learn more about this data stack and about Panoply.io in general, visit <https://panoply.io>

Author: James Green, vExpert

Panoply.io provides end-to-end data management-as-a-service. Its unique self-optimizing architecture utilizes machine learning and natural language processing (NLP) to model and streamline the data journey from source to analysis, reducing the time from data to value as close as possible to none.

© 2017 by Panoply Ltd. All rights reserved. All Panoply Ltd. products and services mentioned herein, as well as their respective logos, are trademarked or registered trademarks of Panoply Ltd. All other product and service names mentioned are the trademarks of their respective companies. These materials are subject to change without notice. These materials and the data contained are provided by Panoply Ltd. and its clients, partners and suppliers for informational purposes only, without representation or warranty of any kind, and Panoply Ltd. shall not be liable for errors or omissions in this document, which is meant for public promotional purposes.