ENTERPRISE BACKUP TO AWS:

A Playbook for Cost-Effective Implementation

Igneous

aws

TABLE OF CONTENTS

INTRODUCTION:
Overview of the data landscape

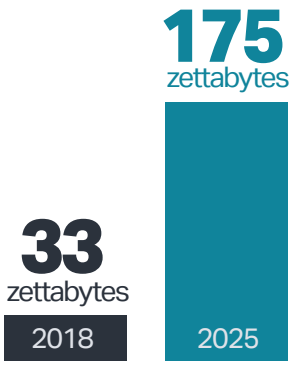
Enterprise data has exploded exponentially over the past decade. Machine-generated data, high resolution imagery and video, internet of things sensors and the data required to feed AI and machine learning algorithms is requiring a massive increase in storage capacity for organizations around the world, both large and small.

The challenge is that approximately 80 percent of this data is unstructured. While structured data often lives in well organized databases, unstructured data like images, videos, documents and machine-generated data often sits in billions of files, mostly in on-premises network-attached storage (NAS) systems. Those files must be organized, moved, managed and protected to meet business needs.

of data is unstructured, in billions of files on on-premises NAS

Most enterprise IT organizations now manage more than 1 billion files and many manage petabytes of file data. And 70 percent of respondents to our data survey in December 2019 say managing unstructured data is difficult with today's tools.

This unstructured data growth shows no sign of slowing down. Analyst firm IDC predicts that the "global datasphere" – encompassing data from datacenters, edge and endpoints – will grow from 33 zettabytes in 2018 to 175 zettabytes by 2025, much of which will be unstructured data, according to IDC's "Modernize Unstructured Data Management" report.

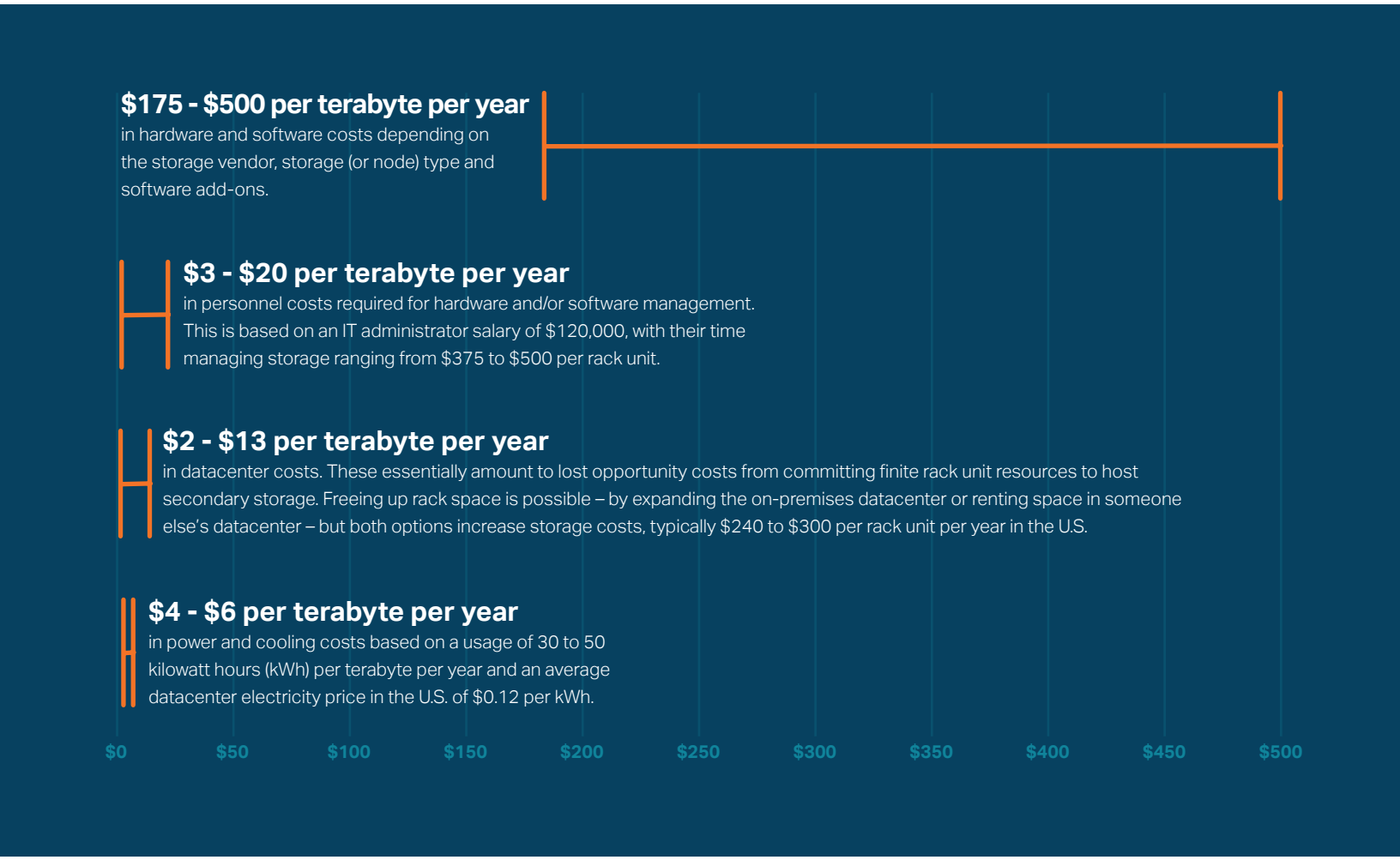


Data growth may increase more than **5x** in **7 years**

All this data poses new challenges, including how to effectively and affordably back it all up. But there is good news. During the past 18 months, data backup on Amazon Web Services has become significantly cheaper than most on-prem solutions - currently costing $12 per terabyte per year.[3]

Companies can realize huge savings from making the switch from on-premises storage to cloud storage for their backup data, savings that weren't possible even 12 months ago.

There are a lot of hidden costs involved in keeping data on-premises vs. moving it to cloud. Companies can easily overlook some of these costs initially and underestimate the expense of on-premises as a backup destination. A more comprehensive review of on-premises costs include:



This brings the total cost of on-premises storage to $184 to $539 per terabyte per year. Backup strategies may rely on disk-to-disk replication to a lower cost storage tier or even tape, but we've run the numbers - on-premises backup is expensive regardless of your approach once you factor in the full costs outlined above. For a comprehensive analysis of the costs of storing data on-premises vs. in the cloud, download our whitepaper The True Cost of Cold Data.

In this eBook, we'll explore why now is the perfect time to weigh alternatives to traditional on-premises backup, including both the key factors that are driving enterprise investment in cloud as a backup destination and six key requirements to look for in a cloud backup solution.

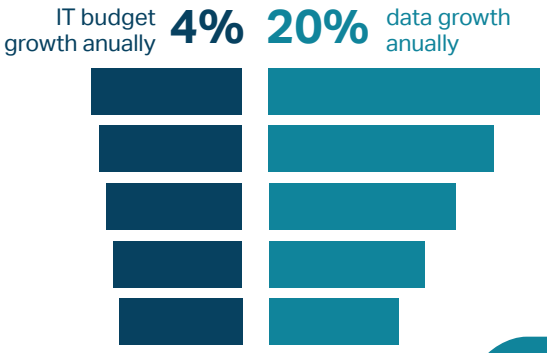
THREE FACTORS

driving enterprise backup to cloud adoption

Moving enterprise backup from on-premises to cloud is increasingly appealing to companies. But why now? Three major drivers are propelling today's growing interest in a cloud backup investment – and are solid reasons to consider doing so within your organization.

Driver #1 – Legacy storage methods can't handle the volume of modern data

Designed in an era of small data, legacy storage methods are inadequate to handle the explosion of data generated today. Characterized industry-wide as "big data," the amount of data would be more accurately described as massive data. In the emerging data economy, where most data is machine-generated, IT departments are having trouble keeping up.



Because of the volume and density of all this enterprise data, legacy storage solutions fail to work effectively in most modern environments. These legacy solutions are also operationally intensive. When overseeing an on-premises data center, IT is tasked with managing server rooms, and all the hardware, infrastructure and overhead associated with them. That's particularly a problem given the discrepancy between IT budget growth and data growth.

Sixty-three percent of organizations say data is growing at a rate of 20 percent or more annually while IT budgets are growing by just 4 percent a year on average, according to research by the Enterprise Strategy Group (ESG), an IT analyst and strategy company.

So somehow they have to deal with this data growth that causes all sorts of other infrastructure challenges. Hardware costs go up. Staffing costs go up. Data protection costs go up. All these things increase as capacity increases if you're still using these traditional storage paradigms and that's creating a challenge.

- Scott Sinclair, Senior Analyst, ESG

OTHER PROBLEMS WITH LEGACY STORAGE SOLUTIONS INCLUDE THE FOLLOWING:

MOST ARE VENDOR-SPECIFIC SOLUTIONS.
That requires standardization on one vendor, which isn't an optimal choice for many organizations with a diverse set of storage issues they're aiming to address.

SOME ARE LIMITED BY FILE VOLUME.
These density limitations mean that millions and billions of files will break most existing backup solutions.

SOME ARE LIMITED BY FILE SIZE.
Legacy solutions often struggle to back up hundreds of terabytes of data and most fail to back up petabytes of data.

MOST REQUIRE AN OFF-SITE COMPONENT.
This isn't always conducive to fast restore (a frequent requirement from companies). That off-site component could involve storing data in a secondary data center or in an off-site tape vault.

Driver #2 – Multi-site, multi-vendor and multi-cloud environments require more flexible backup solutions

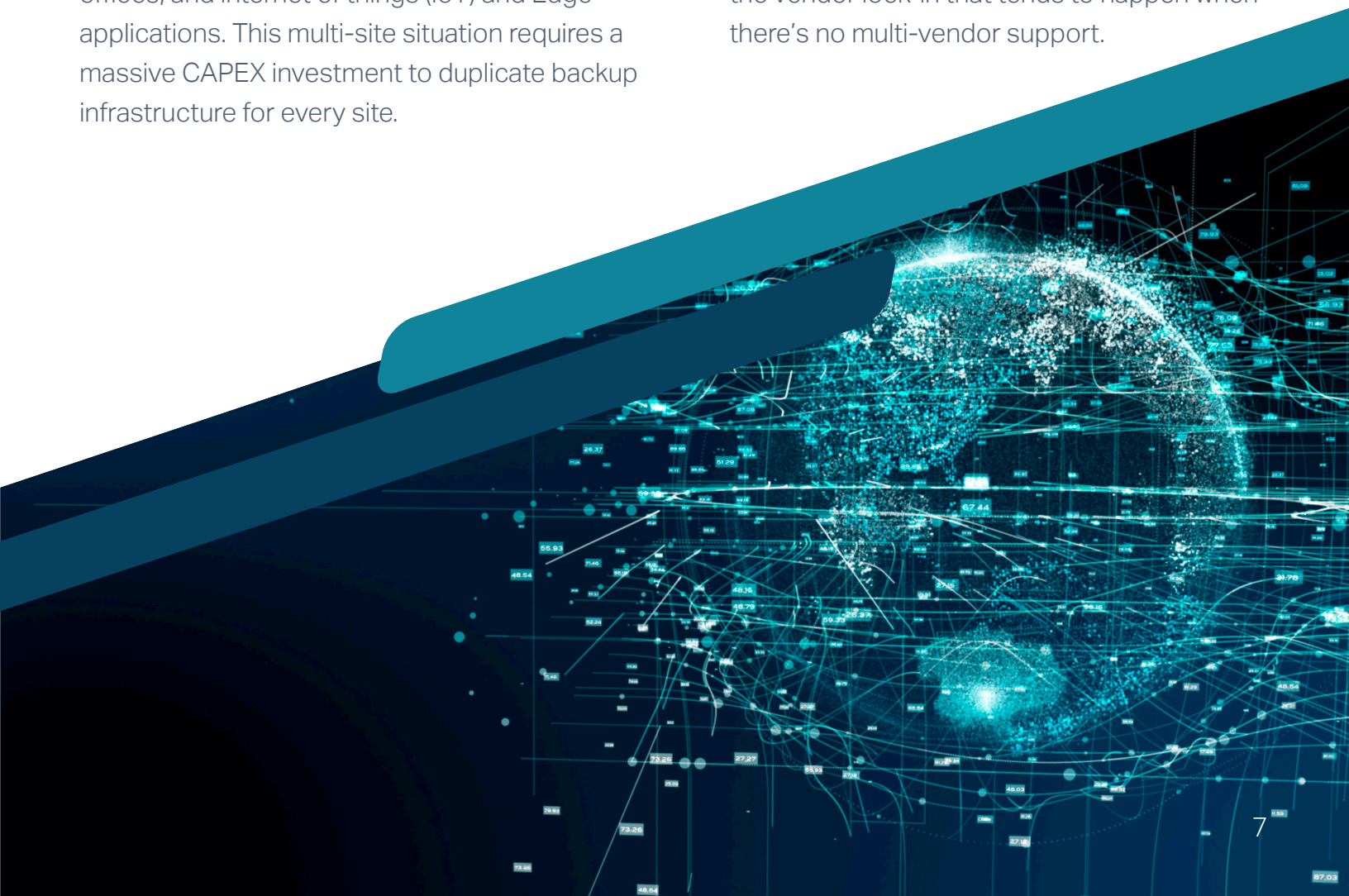
The way enterprises create and consume data has changed dramatically over the last decade.

Many firms now use multiple vendors in their storage stack to meet business needs – a mix of legacy storage solutions and cutting edge, high-performance solutions. Hardware vendors are eager to offer firms backup solutions for their hardware, but that creates complexity as firms bring on new vendors to meet their business needs and they have to support a patchwork of data protection solutions for each hardware vendor.

Today's firms must also support multi-site configurations - Remote sites generate massive amounts of data, which come from research labs, remote development/design centers, field offices, and internet of things (IoT) and Edge applications. This multi-site situation requires a massive CAPEX investment to duplicate backup infrastructure for every site.

Businesses must be able to cost-effectively back up – and quickly restore – data across multiple sites. Serious consequences could result from maintaining the status quo of data storage within the organization:

- In this environment, backup service level agreements (SLAs) can easily be missed.
- With legacy systems, it's cumbersome and time-intensive to get data back, especially when it's stored in off-site tape repositories. Critical data can also be lost because of the lack of visibility.
- Legacy solutions often lack auto-discovery and flexibility. The former problem can mean not knowing if new shares and endpoints are protected. The latter problem results from the vendor lock-in that tends to happen when there's no multi-vendor support.



Driver #3 – Significant cost savings

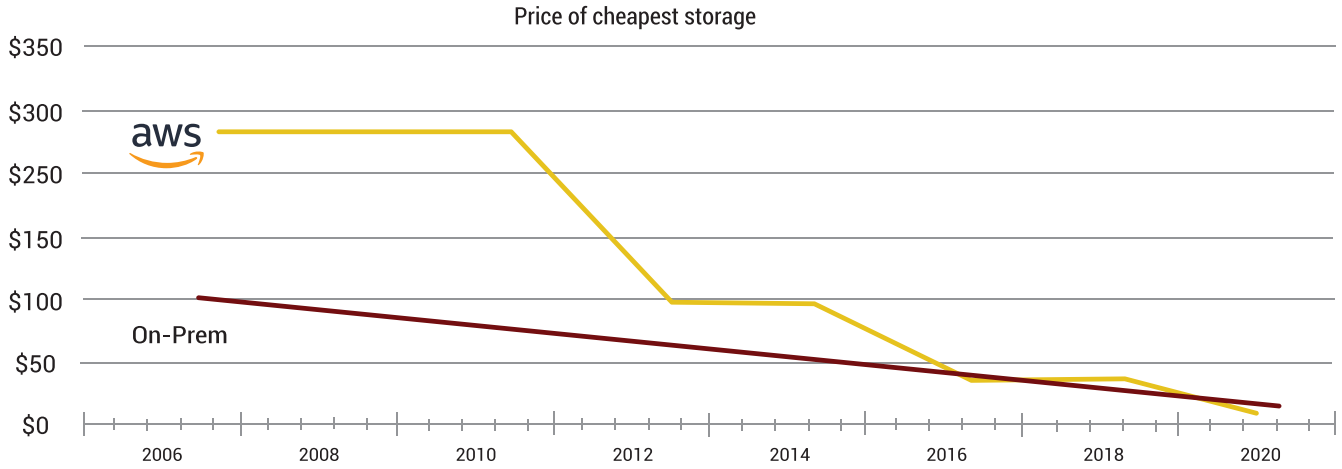
Earlier, we walked through the multiple costs involved with storing backup data on-premises vs. moving that backup data to cloud. A few years ago, an ROI and feasibility analysis of enterprise backup options would have found the cost of on-premises backup fairly comparable with the cost of cloud backup, but those comparisons are no longer accurate.

On-premises storage has become more complex, with the typical customer managing storage solutions from a variety of vendors on a mix of cutting edge and legacy hardware. These multi-vendor, multi-generational storage solutions often don't play well together - creating additional costs in the form of management and maintenance.

On the other end of the on-premises vs. cloud equation is the plunging cost of cloud backup storage. On-premises hardware vendors are trying their best to compete by heavily subsidizing hardware to lock in maintenance and service contracts, but they simply can't compete once maintenance costs, service fees, infrastructure costs and labor is factored in.

With cloud costs for Amazon Web Services (AWS) dropping to unheard-of levels, major savings can be gained from moving enterprise backup to cloud.

Cost of Cloud Storage Over Time ($ per TB/year)



Cloud storage costs are likely to continue their decline which means customers will benefit from future price drops as they happen. Lower prices combined with new software-as-a-service (SaaS) solutions is enabling companies to leverage cloud as a cost-effective backup destination, which wasn't feasible before these recent price drops.

A SaaS data management solution maximizes flexibility by eliminating vendor lock-in. Storage offered by infrastructure-as-a-service (IaaS) cloud providers like Amazon Web Services and a SaaS-based backup tool shifts spend to OPEX and future-proofs your data protection strategy. You can then adopt on-prem solutions that boost performance without having to utilize those same high-cost platforms for data protection.

SIX BEST PRACTICES

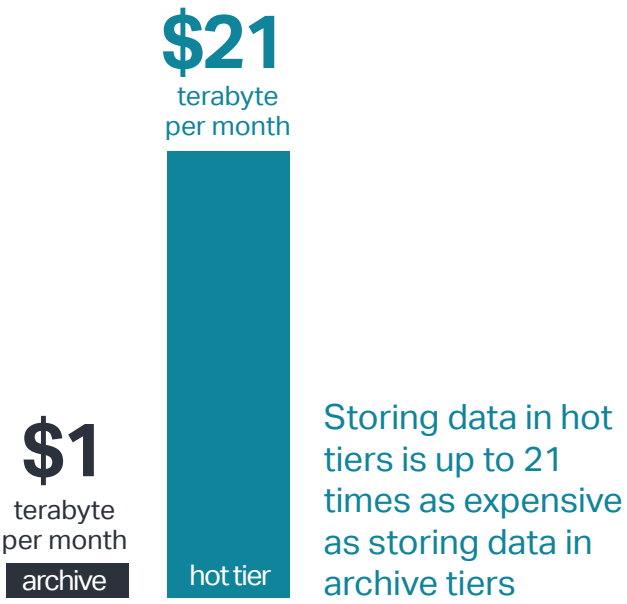
for successful enterprise backup to cloud

As the public and private cloud industry matures, reducing risk and increasing accessibility, more and more companies have been embracing cloud computing to increase their agility and decrease costs while scaling up their business. As we've just explored, businesses have solid reasons for considering enterprise backup to the cloud as part of their overall cloud strategy. However, before transitioning to modern cloud storage there are six best practices you should adopt to ensure success. Get these six things right for a smoother transition.

Best practice #1 – Write data directly to archive tiers

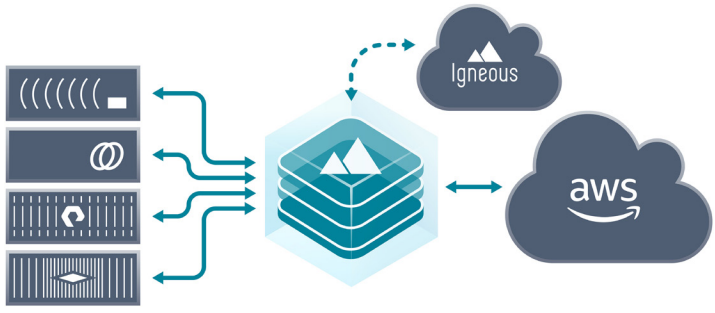
For most companies, backup data acts the same way that car or home insurance does for individuals. It's much better to have it and not need it than need it and not have it. Like insurance, backup offers welcome reassurance that you'll be covered if a worst-case scenario occurs.

But, for many companies, it does not make financial sense to pay for your backup data to have a short recovery time objective (RTO). Unfortunately, many NAS backup solutions will land data only in S3/IA in the AWS cloud. It's up to enterprises to write policies that push the data to lower-cost archive tiers.



Storing data in hot tiers is up to 21 times as expensive as storing data in archive tiers

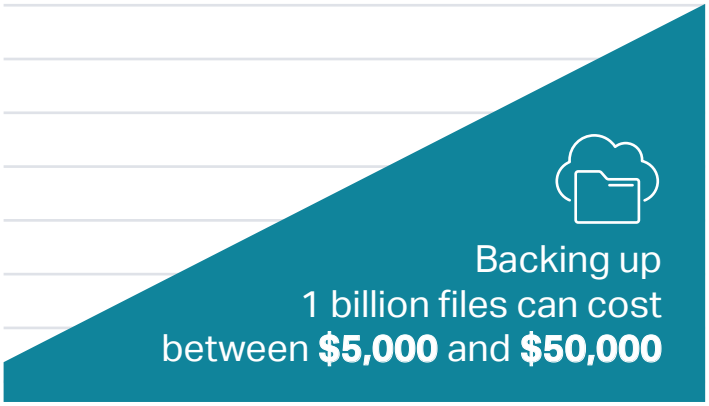
The cost of not pushing data to archive tiers is the difference between around $21 per terabyte per month to store data in S3 Standard, and $1 per terabyte per month to store the data in S3 Glacier Deep Archive. If you write the data first to S3 Standard, then move it to S3 Glacier Deep Archive tier later, that can incur additional transaction costs, which can add up over time.



Instead, opt for a backup solution that natively leverages archive tiers to keep storage costs low and avoid transaction costs of later moving data between tiers. Igneous DataProtect, for example, can target any tier in AWS Simple Storage Service (S3).

Best practice #2 – Minimize transaction costs

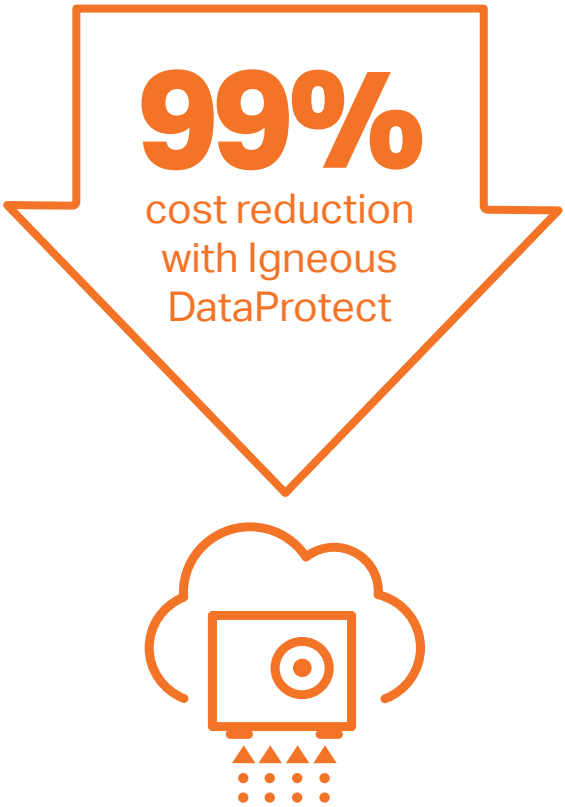
Many vendors promise cloud compatibility but don't optimize for – or even consider – transaction costs. This is an unfortunate mistake when moving NAS data to a cloud tier because these costs (called PUTs) can add up as the number of transactions grows. With costs of between $0.005 and $0.05 per 1,000 transactions, individually backing up 1 billion files to the cloud can cost between $5,000 and $50,000 in transaction fees alone.



Legacy, on-prem disk-to-disk or disk-to-tape solutions don't have to take such costs into account when backing up data. Direct-to-cloud solutions should minimize these costs by design. Make sure you ask your cloud backup vendor to factor transaction costs into the total cost of ownership (TCO) for your solution so you can verify they optimize around this expense.

Igneous DataProtect – with our proprietary IntelliMove technology – operates in a full-once, incremental forever mode. It partitions files into pools based on size before moving data to the cloud. Transaction costs are reduced by orders of magnitude because of chunking, which is when IntelliMove compacts small files (20MB or less) into a larger compressed file.

As a result, Igneous customers can realize a 99% decrease in transaction costs per backup. For 1 billion files, that can reduce transaction costs from a range of $5,000-$50,000 for unoptimized transport to $50-$500 with DataProtect, depending on the public cloud provider and destination tier.



Best practice #3 – Intelligently expire data

Deciding when to expire data often requires careful consideration of legal and financial requirements outside of your control. Such requirements specify how long you must retain data and when you're allowed to expire it. These requirements can include the following:

- Government regulations like SEC17a-4
- Business governance policies
- Service-level agreements (SLAs) with your end users

Expiration policies have to be enforced regardless of where the data is stored. However, cloud providers often set mandatory retention periods dictating how long data must be kept in one of their archive-class storage offerings before they can be expired. For the AWS Glacier Deep Archive, the minimum is three months. Deleting data before meeting these thresholds could result in costly penalties from the cloud provider.

Before putting your NAS backups in the cloud, consider how your data retention policy will be enforced or whether you'll be subject to penalties for deleting data too soon.

Best practice #4 – Know when to clean up expired data

While data expiration involves business policies, reclaiming space in your cloud storage is about optimizing costs. Once you've solved the compliance problem, the data still needs to be deleted from these archive tiers after expiration. To be effective, a storage solution involving archive tiers should operate with business logic stipulating when to reclaim capacity (so that capacity doesn't grow unchecked). Data shouldn't be automatically deleted after expiration.

DATA SHOULDN'T BE AUTOMATICALLY DELETED AFTER EXPIRATION

Look for a backup solution that's intelligent about reclaiming space for backup data. This is in part why Igneous chunks data. It enables us to strike a balance between minimizing transaction costs and managing data expiration.

Best practice #5 – Restore data cost-effectively

The ability to easily restore data provides welcome reassurance that you can get your data back when you need it. Most restore operations initiated by business users involve directories or individual files. There are a few key questions to ask when considering a modern solution, which are as follows:

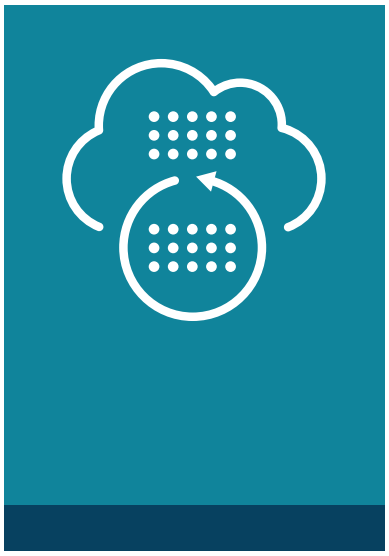
- Can you restore just what's requested versus significantly more than that? When small datasets need to be restored, the process should ideally be efficient and restore only what's needed. If you have to do a large scale restore to recover a small number of files, you'll incur unnecessary transaction costs, require more storage on the destination system than needed and it may even slow down the restore process due to network speeds in moving the data.

- Does the solution offer the option of an immediate restore versus a bulk restore based on urgency? A bulk restore usually meets most business's SLAs at a significantly lower cost than an immediate restore does.

- Can it restore data to any storage tier, including both on-premises NAS appliances and file systems in the cloud?

In our experience, restore rates of cloud backups to on-premises NAS hover between 0.25% to 3%. Most restores involve directories, files, and the occasional export/ share that must be reverted.

AN AVERAGE RESTORE FROM CLOUD TO ON-PREM CAN BE AS LITTLE AS 1TB AND UP TO 500TB



3%
of total storage

0.25%
of total storage

Igneous chunking pays off with restore, enabling us to restore only the chunks needed for the recovery, whether they comprise a directory or group of files. This minimizes restore and retrieval costs. On average, a restore will cost $35 to $70 per terabyte out of the cloud to on-prem, or about half that much if a customer restores the data to a cloud file system.

Best practice #6 – Move data quickly and securely to the cloud

Moving data quickly to the cloud requires the use of a direct connection to your preferred cloud storage system. Amazon offers a direct connection to their cloud called AWS Direct Connect. Typically billed per hour, these direct connections are usually available in 1G, 10G, or sometimes 100G options.

A direct connection is generally preferable because it offers a much larger bandwidth connection to the cloud. Igneous DataProtect can move much larger volumes of data per day to the cloud if such a connection is in place. Customers generating or changing large volumes of data per day will require this extra connectivity to ensure their backup SLAs are maintained.

HERE'S HOW THE AWS DIRECT CONNECT WORKS:

AWS S3 storage is accessed using public object interfaces and doesn't have internal interfaces for customers leveraging direct connections. This gap is filled using AWS Proxy Server for the data to move through. Customers can then direct specific applications or services with high reliability and security to route traffic through this connection while leveraging the public internet for less sensitive applications.

Consider a direct connection for fast, secure backup to the cloud.

IGNEOUS: A MODERN SOLUTION FOR ENTERPRISE-GRADE BACKUP TO CLOUD

Motivated by the many benefits of backing up data to the cloud, companies are choosing Igneous for secondary storage in the cloud as well as for management of unstructured data that must remain on-prem. Two-thirds of unstructured data has moderate to extreme value, according to the respondents to an Igneous data survey. Data that is critical to your business must be protected.

MOST UNSTRUCTURED DATA HAS MODERATE/ EXTREME VALUE

Built for speed and scale, Igneous DataDiscover and Igneous DataProtect deliver the high performance needed for modern enterprises. Whether on-prem or in the cloud, data can be managed anywhere it lives and moved anywhere it needs to be.

Any NAS environment

Flexibility is required for any modern data management solution. The flexibility offered by Igneous means you can analyze and manage your data across all NAS devices, including multi-vendor and multi-site environments through a single interface. We support ultra-fast flash-based storage like Pure FlashBlade, modern file systems like Qumulo File System, and industry leaders like DellEMC Isilon and Netapp and we offer generic SMB and NFS support for any other vendors.

Your data protection solutions should be independent of your vendor alignment for storage capacity - this approach gives you maximum flexibility in choosing the right NAS partner to meet your business needs while ensuring a single interface to manage data protection across your storage infrastructure. Your NAS provider shouldn't dictate your data protection solution.

Any storage destination

While support for any NAS data source is critical to unify data management across storage hardware on-premises, support for any on-premises or cloud tier is equally important. At Igneous, we are increasingly seeing companies store backup and archive datasets on a mix of on-premises hardware and cloud targets in order to reduce risk, improve performance, and optimize costs.

While Igneous can target any tier in AWS Simple Storage Service (S3), we also support on-premises hardware as your backup or archive destination. If your business needs warrant it, you can write your backup or archive data sets to multiple destinations - either in fully on-premises configurations, cloud-only configurations or hybrid cloud configurations. Your data protection tool shouldn't dictate where your backup and archive data resides.

RESEARCH ORGANIZATION CHOOSES "CLOUD AGNOSTIC" IGNEOUS

Using machine learning (ML), research organization **PAIGE** helps pathologists diagnose diseases more efficiently. Training the machine-learning algorithms requires petabytes of high-resolution tissue scan images. Just 40 of these slides are larger than most datasets in the world. PAIGE chose Igneous because of its scalable data protection, seamless data movement, easy data discovery process, and cloud compatibility.

"One of the main reasons that we chose to go with Igneous is because they are cloud agnostic," said Ran Godrich, research engineer at PAIGE. "As a young company, we always want to use the latest offer packages that are coming out every week, every month, in the world of artificial intelligence and machine learning. We don't want to limit ourselves to one. And what Igneous allows us to do is choose whatever software you want, whatever cloud provider you want, and whatever machine learning models you want to build. And that's really what we're looking for.

- Ran Godrich, Research Engineer, PAIGE

Zero hardware, Cloud Native and Delivered as-a-Service

Because Igneous unstructured data management is an as-a-service solution, there's no hardware for IT to manage. That frees your organization of three of the biggest burdens of data management. It means no infrastructure management, no software management, and no troubleshooting.

Many products are marketed as cloud native, but customers are regularly asked to upgrade to the latest release. Igneous unstructured data management is truly cloud native, meaning you always have our latest code and never have to file change requests to get to the next version. You don't have to ask vendors to back-port bug fixes to your version or figure out how to upgrade five or 10 sites. Bottom line: true cloud native means easy management.

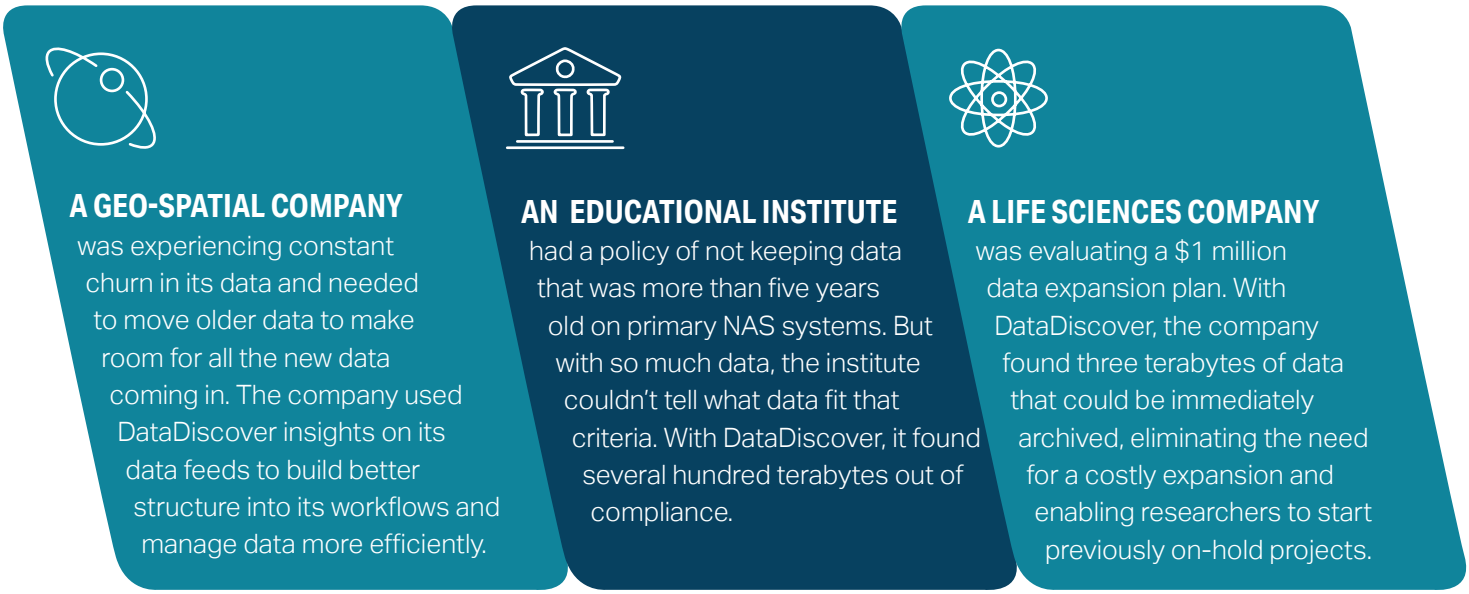
Through our as-a-service solution, Igneous proactively monitors customers' data protection workloads, alerting you if your backup or archive jobs encounter a problem. Our proactive monitoring and customer communications set us apart, offering peace of mind to companies like Quantum Spatial, a geospatial solutions provider in North America. The company sought help from Igneous to manage petabytes of geospatial data across multiple locations.

> **Igneous monitors the system and lets me know if anything is happening. It helps me sleep at night.**
>
> - Travis Spurley, Senior Systems Engineer at Quantum Spatial [view the client story]

Multi-site visibility

Igneous DataDiscover lets organizations understand what data they have, where it lives, and how long it's been in their environment. This comprehensive view of data includes the location of the 60 percent of their data that's cold or inactive and could be a target for a data archiving strategy to further reduce primary storage needs. This solution takes just minutes and is the easiest way to analyze petabytes of file data across multiple locations and NAS systems.

When you have hundreds of millions of files, what you don't know can cost you. It's a valuable lesson learned by three companies:



A GEO-SPATIAL COMPANY was experiencing constant churn in its data and needed to move older data to make room for all the new data coming in. The company used DataDiscover insights on its data feeds to build better structure into its workflows and manage data more efficiently.

AN EDUCATIONAL INSTITUTE had a policy of not keeping data that was more than five years old on primary NAS systems. But with so much data, the institute couldn't tell what data fit that criteria. With DataDiscover, it found several hundred terabytes out of compliance.

A LIFE SCIENCES COMPANY was evaluating a $1 million data expansion plan. With DataDiscover, the company found three terabytes of data that could be immediately archived, eliminating the need for a costly expansion and enabling researchers to start previously on-hold projects.

Infinitely scalable

No longer does scalability require adding physical servers. Igneous offers infinite scalability by automatically adding or destroying microservice instances as needed.

Altius Institute for Biomedical Sciences generates enormous amounts of data from editing DNA. One of its objectives is to see if changing the structure of the nucleus can change the ways cells behave and stop the spread of cancer. Using legacy systems to manage more than two petabytes of data wasn't sustainable. Igneous' scalability was very appealing, giving Altius Institute room to grow.

> **In the past, as the amount of data increased, it would require me to write a bigger check to address the issue. But now, because of Igneous, I don't have to hire any more people, I don't have to take on any more risk, and I don't have to increase my management in any way.**
>
> - Michael Cockrill, Chief Technology Officer, Altius

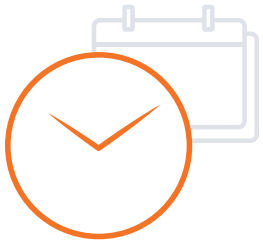
Cost efficient

By helping companies move more data to archives tiers and reclaim NAS capacity, Igneous unstructured data management can reduce storage costs by up to 50 percent.

An Igneous customer on the east coast had generated so much unstructured data – 32 billion files comprising 400 terabytes of data – that it had trouble backing it all up. Complicating things, the company needed a complete NAS refresh, but its infrastructure vendor said it couldn't do that until a complete data backup had occurred. Alternatives would have taken months to deploy and cost hundreds of thousands of dollars. Instead, the company chose Igneous, which deployed within hours and at a cost that was less than a third of what the alternatives cost.

Igneous

Hours instead of months



Less than 1/3 the cost

Every unused terabyte of data that is discovered and archived using DataProtect could save $195 to $2,028 per terabyte per year, according to "The True Cost of Cold Data" report. Use our cost calculator to estimate how much your organization could save by archiving unused data or contact us if you'd like a custom TCO analysis for your specific environment.

Fast deploy

Companies can deploy Igneous as-a-service solutions in hours, not months.

Because Igneous is delivered 100% via software and deployed via a virtual machine in your environment, installation is not a complex IT project. Deploy our stateless VM in minutes, import any NAS system in seconds, and view results in minutes to hours. Because we are deployed as-a-service, you can request a trial of our product and get started immediately - no site visits, no hardware deliveries, no data center footprint required.

Deployed in minutes
100% via software

Lightning fast

In addition to a speedy deployment, ongoing management is lightning fast with Igneous. That's true whether you protect file data onsite, move secondary storage to the cloud, or employ a hybrid cloud setup. Igneous DataProtect, for example, is the fastest way to backup petabytes of data and billions of files, reducing IT team hours by up to 90 percent.

Igneous makes it much easier for IT to manage data and reduces IT worries.

Next steps: Modernize Your Data Management in 90 Days or Less

Igneous is optimized for deploying rapid proof-of-concept environments to our customers. You can request a POC and deploy our VMs in your environment in minutes or you can sign up directly through our AWS Marketplace listing. We can assist you with the setup, or you can do it on your own. Whether you are currently managing a multi-vendor, multi-site, multi-petabyte environment or are looking to push your first terabyte from on-premises to the cloud, we can help.

CONTACT US

Already an AWS customer? Find us on:

aws marketplace

GET IN TOUCH

2401 4th Ave
Suite 200
Seattle, WA 98121

1-844-IGNEOUS
or 206-504-3685

info@igneous.io

Igneous