druva

White Paper

IT UPDATE: LEVERAGING THE CLOUD TO SOLVE TODAY'S ARCHIVING CHALLENGES

How to Leverage the Cloud to Easily Address the Global Regulatory Environment

druva

Consider This First

Before you submit that new bill of materials or hardware requisition order for more storage, consider this. Archiving goes beyond just long-term retention—it is designed to help organizations meet their legal, compliance and data access needs by managing analytics, search and legal hold functions across underlying data sets. As cloud technologies mature and undergo widespread adoption, more and more companies are moving data archiving to the cloud. Incentivized by simplicity, massive cost savings, more timely access to data, and technology trends that make legacy systems obsolete, organizations are exploring cloud archiving like never before and searching for best-of-breed technologies to get the job done.

However, cloud archiving is still in its nascent stage and many CIOs and other IT leaders, only vaguely understand its value proposition above and beyond its ability to reduce costs. This paper is designed to help demystify today's cloud archiving environment and educate leaders about key issues to consider before they move to the cloud.
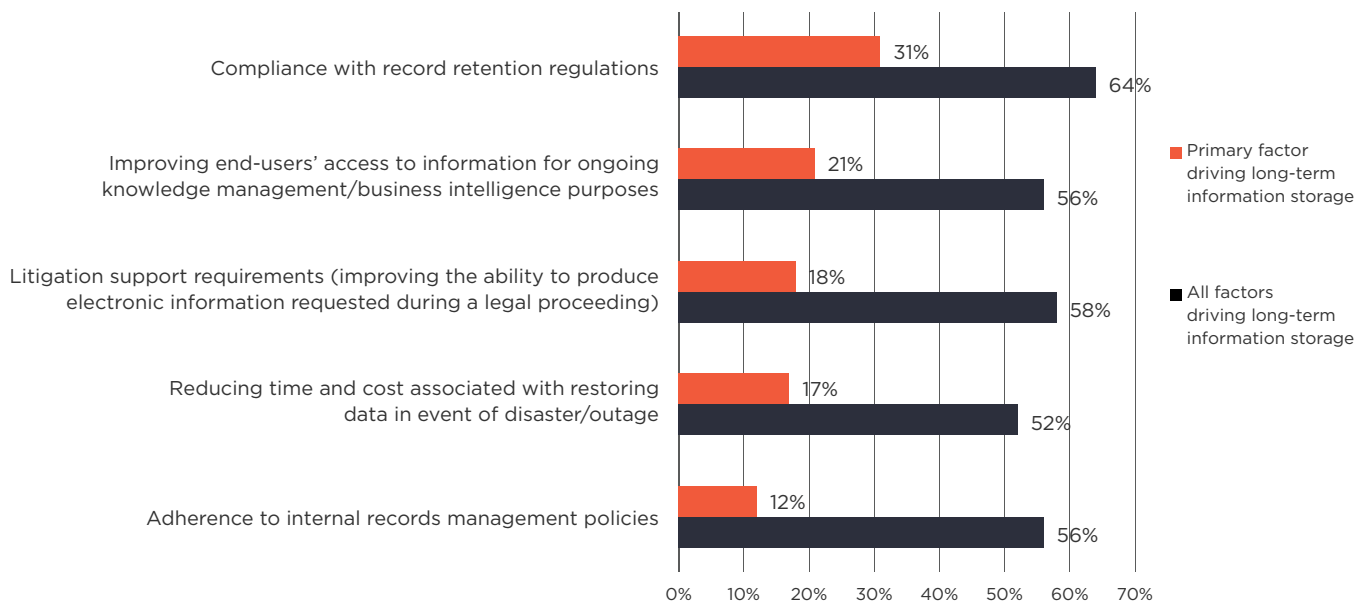
Where Cloud Archiving Adds True Value

- **Regulatory and internal compliance —** Enabling the business to cost-effectively store data over time to align with retention, auditability and accessibility requirements.
- **Legal and eDiscovery support —** Ability to search, preserve and maintain chain of custody of data over the long term for investigative and litigation needs.
- **Application retirement and long-term data retention —** Maintaining long-term storage and access of application data to meet retention and knowledge needs of the business.

Technology Trends Alter the Archiving Landscape

The emergence of cloud archiving—and corresponding demise of long-used legacy solutions—isn't random; it's driven by technology trends that have irrevocably changed how organizations retain data for legal and compliance purposes.

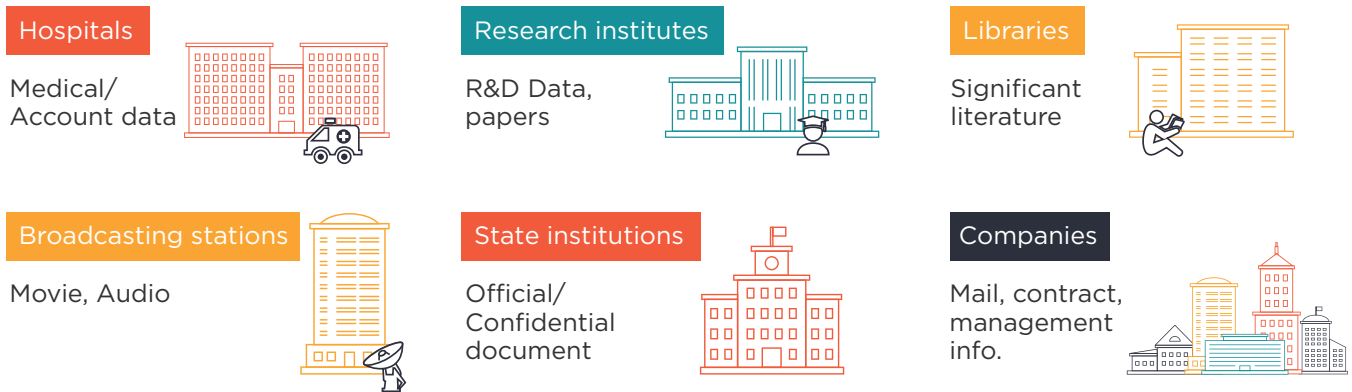**Reasons for Storing Information for Long-term Retention and Reference**

Compliance with record retention regulations — 31% / 64%

Improving end-users' access to information for ongoing knowledge management/business intelligence purposes — 21% / 56%

Litigation support requirements (improving the ability to produce electronic information requested during a legal proceeding) — 18% / 58%

Reducing time and cost associated with restoring data in event of disaster/outage — 17% / 52%

Adherence to internal records management policies — 12% / 56%

Legend:
- Primary factor driving long-term information storage
- All factors driving long-term information storage

X-axis: 0% 10% 20% 30% 40% 50% 60% 70%

Increased Regulation

In the wake of various corporate scandals and the financial crisis of 2008, increased and more stringent regulations in the U.S. and abroad require companies to more diligently archive, store, and retain data for legal and compliance needs. One of countless global regulations covering data retention periods, the Sarbanes-Oxley Act (SOX), instituted after the gross financial fraud of companies like Enron in 2002, requires that work papers, memos, emails and other documents that form the basis of an audit or review be retained for a period of 7 years. Pending litigation is another area where data must be held for indefinite periods of time. Incidents where data was improperly kept, like in the precedent-setting Zubulake vs. UBS Warburg, further show that companies may have to bear significant financial burdens for not fulfilling eDiscovery requests.

The upshot of increased regulation is that companies are under ever-increasing pressure to hold data for compliance audits, litigation, and other legal uses. The odds of such events are high: according to the Norton Rose Fullbright 2015 Annual Litigation Trends Survey, 37% of large organizations reported having 20 or more pending lawsuits. The Survey further shows that 42% of companies overall reported more than 6 lawsuits filed within the last 12 months, with cases exceeding $20 million reported by 26% of those surveyed. Not surprisingly, the ESG research report Backup and Archiving Convergence Trends finds that regulatory compliance tops the list of reasons that organizations store data on a long-term basis.

Archiving "industry-classified data"

Hospitals
Medical/Account data

Research institutes
R&D Data, papers

Libraries
Significant literature

Broadcasting stations
Movie, Audio

State institutions
Official/Confidential document

Companies
Mail, contract, management info.

Preservation terms of digital documents defined in laws

Terms	Target documents	Related laws and regulations
30 years	Patient Information, Electronic Health Chart, Medical Records	USA: HIPPA/OSHA
15 years	Design Documents	Licensed Architect Act
10 years	Brokerage: Trade Reports Manufacturing Drawing (10 years after end-of-product)	Cabinet Office Ordinance on Securities Product Liability Act
7 years	Journal, Account Book (General Ledger) eDiscovery	Income Tax Law, Corporation Tax Law USA: Sarbanes-Oxley Act
5 years	Medical: Medical Records, Midwifery Records, Emergency Medical Records Corporate: Property formation tax exemption savings application	Medical Act etc.Income Tax Law

Explosion of Enterprise Data

The growth of enterprise data is another trend putting pressure on companies to more efficiently archive, retrieve and manage information. As noted in an article on CIO.com, an IDC study finds that most businesses experience between 40–60% growth in data volume annually. The distributed enterprise, rise of mobile computing, and proliferation of applications, platforms and networks are all key drivers of growth, which is exacerbated by long data retention periods. In this computing environment, it's easy to see why companies find it increasingly difficult to store, manage and access data using yesterday's technologies.

Archiving and Backup Are Complementary

There's a growing awareness that archiving and backup are related operations that require complementary solutions in order to achieve optimal results. Whereas backup is used for recovery of missing or corrupted data from a particular point in time, archiving represents the long-term retention and management of infrequently-accessed data. Despite this distinction, many organizations still use traditional backup software, rather than an archiving solution, for long-term retention. This can lead to considerable expense due to the way that most backup software stores data over time, whereas archiving solutions are optimized to store data more efficiently and provide compliance related capabilities like search on the underlying data set. At the same time, there are organizations that continue to handle archiving and backup as completely isolated processes, requiring separate technology solutions, data protection silos, and management, basically doubling the overall expense. As the problems with both approaches continue to surface, companies that want to remain competitive are adopting new solutions that integrate and automate what used to be two separate, siloed functions.

Legacy Approaches Recipe for Failure

Data archiving is not optional, but represents a mission-critical enterprise operation. Despite this, legacy solutions have proven woefully inadequate at meeting enterprise archiving needs, especially given the new data landscape. Traditional methods employing a mix of on-premise infrastructure, disk hardware and tapes are extremely costly and labor-intensive given the long-term nature of archive storage—something easily overcome by leveraging cloud instead. The complex workflows and multiple vendors required by legacy solutions further add to their inefficiency and expense.

As detailed below, legacy archiving solutions face a number of common storage challenges:

Disk

Disk is still the predominant media type for storing secondary and tertiary data. In fact, ESG reports that on average, more than 50% of archived data volumes are retained on hard disk drive technology. However, disk is an extremely expensive medium for storage at the required scale, leaving many organizations unable to afford the redundant infrastructure necessary to ensure data integrity and resiliency. This substandard approach to archiving leaves organizations vulnerable to lost and corrupted data, non-compliance and other issues.

Tape

Like disk, tape often fails to meet compliance and other data governance needs. Due to long-term retention policies, a growing number of systems and sites, and a wildly impractical "hold-everything-just-in-case" mindset, companies often manage thousands, or even hundreds of thousands, of aged tapes in libraries and offsite locations. At this scale, performing tape-based backup and manually managing and shipping tapes from remote offices to central repositories is time—consuming and inefficient—in effect, the more sites, the more complex, costly and inefficient it becomes. Retrieving and restoring legacy data from mountains of tape is likewise slow and cumbersome, a major burden for companies facing time-sensitive compliance or legal demands. In addition, boundless data retention can lead to serious legal exposure if information is asked for during litigation and no solution exists to automate deletion based on standard retention policies.

A number of technology issues can also render tape unrestorable. Hardware, software, configurations and administrator skill sets need to match the environment in which the tape was created, a compatibility requirement that becomes increasingly difficult over time. Encrypted tapes subject to recoverable keys can also impede data retrieval if processes and systems become outdated or passwords are misplaced. In addition, tapes are heavily subject to compromised data integrity, with a single tape corruption capable of rendering an entire archive obsolete.

Cloud Archiving Outperforms Alternatives

Unencumbered by the burdens and limitations of legacy solutions, the unique capabilities of the cloud enable enormous cost savings and data management functionality that were never before possible. As noted in the slide below, data backup and archive are now the top cloud infrastructure use case.
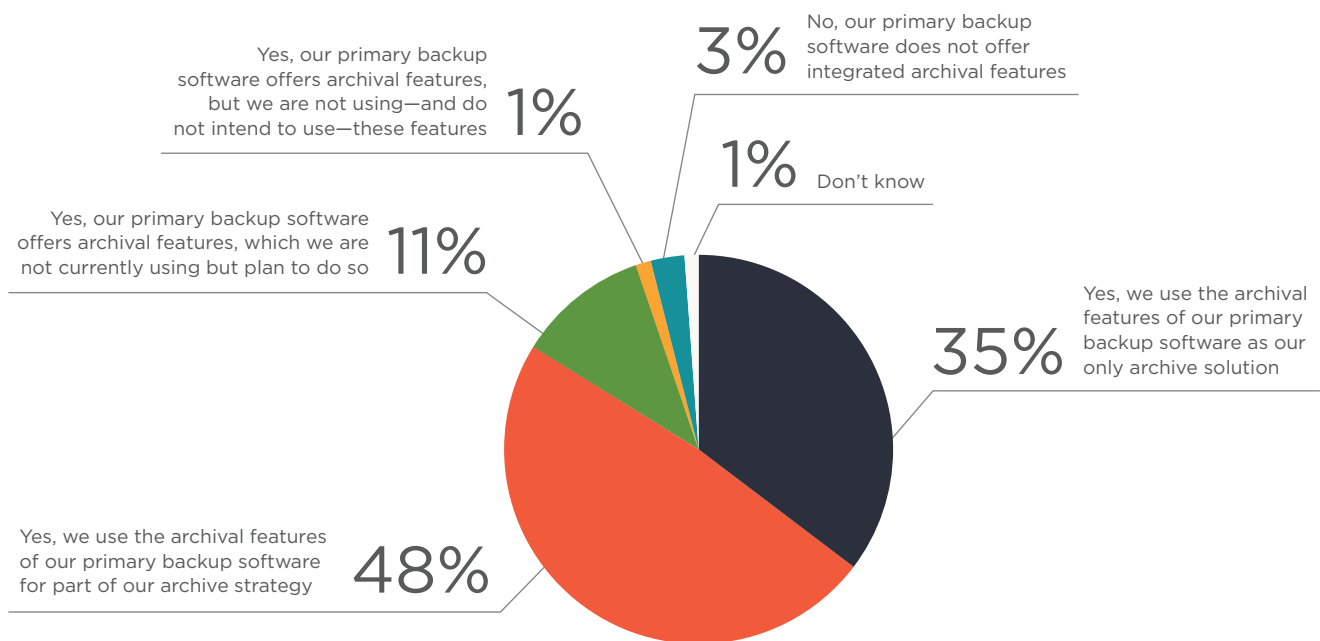
There are numerous reasons why cloud archiving solutions have surpassed legacy systems as the best option for storing and managing this type of data over time. Performing far beyond on-premise systems, such cloud solutions offer:

- **Improved regulatory compliance** with high availability and data governance functionality that spans all system data and leverages cloud elasticity to process data much faster than on-premise systems.
- **Much faster access** than offline tape, enabling legal or compliance teams to do their work on-demand without having to wait.
- **Superior security model** due to the heavy investments in security infrastructure. Not surprisingly, regulated businesses requiring top-notch security are adopting leading providers like Amazon Web Services (AWS) 3x faster than non-regulated businesses. Cloud archiving is also devoid of the security challenges inherent in manual processes and physical data assets.
- **Global availability and resilience** across variable devices, networks and locations—a key requirement for distributed companies processing massive volumes of data.
- **Simple, standardized access protocols for long-term access** rather than the proprietary software and hardware required to read data on tape.
- **Vast, scalable storage capacity** that expands dynamically, without advance planning, to meet growing archive requirements.

- **Zero infrastructure requirements,** thereby removing the need for pricey, on-premise facilities, redundant systems and staffing overhead, and enabling companies to eliminate off-site tape vaulting services and otherwise consolidate management sites.
- **Freedom from hardware/software maintenance and upgrading** that can create version issues and incompatibilities in legacy systems and increase administrative complexity when spread across multiple sites.

But Not All Cloud Approaches Are the Same

IT leaders shopping around for cloud archiving technology should keep in mind that not all solutions are alike, with a distinct difference between those storing data for long-term retention and those offering archive storage. Accordingly, there are several things organizations should consider as they assess the best approach for cloud archiving for their business.

- **Gateways vs. Direct to Cloud —** There are different approaches for sending data to the cloud, with gateway appliances sometimes serving as targets for backup and long-term storage. However, such appliances often function as nothing more than "dumb" dumps where data is stripped of useful context. Depending on your system topography, cloud gateways can also add unnecessary complexity and management overhead from a locality perspective. In contrast, direct-to-cloud archiving ensures that data context is maintained, with a unified access and management point provided across all localities.

- **Cloud as a Storage Medium —** It's not enough to simply store data long term in the cloud; directly acquiring cloud storage space and moving data via scripts or direct copy winds up being highly inefficient and costly over time. Data lacks context and becomes complex to manage as years of data are stored. For true archiving, companies require search, fast retrieval and management functionality to attain the insights necessary for legal, compliance and other data governance purposes. To yield the most actionable data possible, it's important to select a cloud solution that provides visibility and control, and can access data at a granular, rather than in a monolithic fashion.

- **Cost Models Can Vary —** While cloud archiving is indeed capable of delivering massive cost savings, pricing can vary based on technology and storage methodology. Absent deduplication or streamlined backup methods and policies, simply changing the backup target from tape to cloud could make storage much more expensive and data difficult to retrieve. Organizations should also be on the lookout for hidden or underestimated costs associated with retrieving archived data. To maximize cost savings, companies should adopt a cost-effective, "pay-as-you-go" pricing model that's aligned with actual, rather than projected, usage.

Druva's Cloud-Native Solution for Efficient, Cost Effective Archiving

By leveraging a unique and scalable cloud native storage architecture, Druva overcomes the outdated and overtaxed legacy archiving models and delivers the search, retrieval, and management functionality and speed required for today's legal and compliance needs. Through its highly efficient data deduplication system, organizations using Druva see unprecedented cost savings and eliminate the inefficiencies and storage bills of legacy archiving solutions.

How Druva Adds Value To Cloud Archiving

Long-term policy based data retention

- Dynamic retention policy management profiles
- Data lifecycle management

Litigation search and discovery

- Global content-aware search capabilities
- Easy access to data during collection
- Direct connections for eDiscovery with third party tools

Regulatory compliance management

- Efficient and continuous immutable data capture
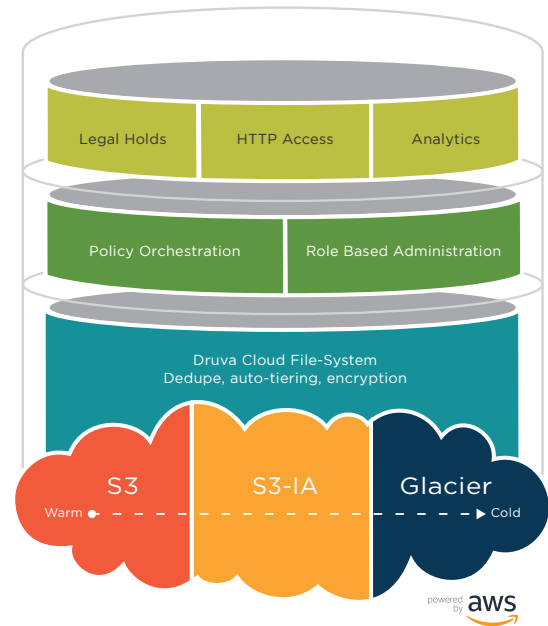- Central visibility into heterogeneous environments

Application retirement automation

- Integration with on-site applications

Cost Efficiencies

- Globally deduplicated data on AWS S3 Object Storage
- Tiered storage on AWS Glacier

Cloud Archiving on AWS

Legal Holds | HTTP Access | Analytics
Policy Orchestration | Role Based Administration
Druva Cloud File-System
Dedupe, auto-tiering, encryption
S3 — S3-IA — Glacier
Warm — Cold
powered by aws

Cloud Data Sources
IaaS | PaaS | SaaS

Local Data Sources
Endpoints | Branch Office | Data Center

druva
powered by aws

Data Protection
Backup & Restore | Disaster Recovery

Data Governance
Archival | Compliance | eDiscovery

Data Intelligence
Machine Learning & Analytics

Key Advantages of Druva's Cloud-Native Approach Include

- **Reduced Infrastructure Burden —** With its vast storage capacity, data resiliency and immediate data accessibility, Druva's archiving approach reduces the need to own or lease hardware infrastructure. As a result, companies that move data from expensive, on-premise storage to cloud archives can reduce both operational costs (OPEX) and capital expenditures (CAPEX). Because it enables centralized data management, its cloud native architecture further reduces IT's reliance on on-premise facilities, hardware and staff.

- **Efficient Data Management and Storage —** A cloud native approach enables organizations to significantly reduce their storage footprint and overhead through a variety of means. First and foremost is global deduplication that deduplicates data across servers and sites, dramatically lowering the enterprise data footprint (up to 20x) and enabling massive bandwidth savings. Ever-incremental archiving, which synchronizes only incremental changes to content, further reduces data volume while providing storage flexibility and simplifies recovery management. Similarly, auto-tiering of data, which leverages the cloud's native data and infrastructure resiliency, reduces storage costs by making warm data available for instant restores while routing infrequently accessed, archival data to less costly storage.

- **Multiple Data Workloads on a Single Set of Data —** Rich search and data management features made possible by cloud native architecture enable organizations to unify multiple workloads and leverage archived data for multiple use cases, eliminating the need for separate, redundant data replication systems. With cloud, companies can archive, categorize, and search datasets in support of eDiscovery, compliance, data analytics, and other business purposes. Such centralized data management maximizes the value of data and reduces separate legacy systems and workflows, leading to increased efficiencies and cost-savings.

- **Payment Models Shrink Total Cost of Ownership (TCO) —** Payment models that charge organizations only for the storage they use eliminate wasted capacity reservations. In stark contrast to the complex pricing formulas of legacy solutions, such models effectively shrink TCO. Unfortunately, organizations routinely overlook the cost savings of repurposed hardware in their TCO calculations, leading to falsely elevated TCO figures.

As companies educate themselves about using cloud for long-term retention needs, it's essential that they first understand the fundamental differences between data archiving and long-term storage. Once this dynamic is fully understood, organizations can rapidly ascend the cloud learning curve and make better, smarter decisions in choosing a cloud storage solution that best achieves their strategic goals. Because of its deep understanding of the unique requirements of data archiving, Druva has adopted a cloud-native approach that meets real-world business needs. As a result, Druva delivers dramatic cost savings hand-in-hand with rich functionality and high performance – so customers don't have to skimp on the essentials. Given this high-powered combination of cloud-enabled features and benefits, legacy archiving solutions—which fall far short for data archiving—are on their way out. Instead, companies considering the transition to cloud for archiving can look to Druva for help so they can stay competitive in a business environment where looking backwards or standing still is simply not an option

What's Your Next Step? Perform a TCO Comparison

Druva enables you to reduce costs by eliminating the need to make additional capital expenditures on underutilized hardware and provides you with a pay-as-you-go model. By reducing the Total Cost of Ownership (TCO) of your archival solution, you're able to invest in only the capacity you need, freeing up much-needed budget for other critical projects that move your business forward.

Performing an effective TCO that empowers the business to make the right decisions, requires detailed and careful analysis. It's important to consider several factors including hardware procurement, management costs, maintenance fees and the time and effort that goes into upgrades, patches and decommissioning. All of these factors will be necessary in order determine the real cost of your on-prem solution over its useful life.

This exercise can be a challenge for most organizations to really get it right, which is why Druva offers both an online TCO comparison tool as well as a personalize detailed TCO consultation in order to assure that all factors are considered within your unique environment.
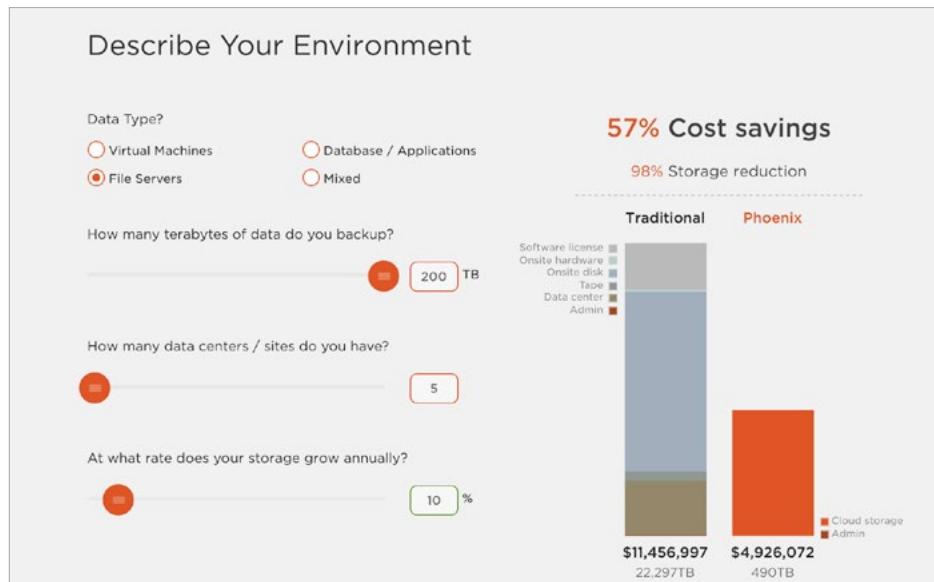
The example:

- Data Type: File servers
- Data Size: 200TB
- Number of data centers: 5
- Data growth (annual): 10%

The assumptions

- 3 year TCO comparisons include costs of legacy hardware, maintenance, upgrades, admins and vendor support
- Server annual maintenance of 15%/yr
- Retention assumptions: 3 year period
- Traditional backup assumes a disk-to-disk-to-tape methodology that includes backup software licenses, disk hardware and tape infrastructure for offsite backup
- Nominal daily data change rate of 1%
- Retention assumes a typical Grandfather/Father/Son (GFS) schedule with weekly and monthly full-backups/snapshots

You can find our TCO calcualtor here: http://www.druva.com/products/phoenix/pricing-model/

Conclusion

With limitless snapshots and infinite retention capabilities, Druva's enterprise trusted, high-performance cloud platform delivers centralized management of hot, warm and cold backups, while removing the need for outdated, error-prone and heavily manual approaches — significantly lowering costs and providing enhanced functionality that delivers value beyond cost savings.

About Druva

Druva is the global leader in Cloud Data Protection and Management, delivering the industry's first data management-as-a-service solution that aggregates data from endpoints, servers and cloud applications and leverages the public cloud to offer a single pane of glass to enable data protection, governance and intelligence—dramatically increasing the availability and visibility of business critical information, while reducing the risk, cost and complexity of managing and protecting it.

Druva's award-winning solutions intelligently collect data, and unify backup, disaster recovery, archival and governance capabilities onto a single, optimized data set. As the industry's fastest growing data protection provider, Druva is trusted by over 4,000 global organizations, and protects over 40 petabytes of data. Learn more at www.druva.com and join the conversation at twitter.com/druvainc.