ACTIVE Archive

ANNUAL REPORT

April 2023

EFFECTIVE DATA MANAGEMENT THROUGH ACTIVE ARCHIVES

The Active Archive Alliance serves as a vendor-neutral, trusted source for providing end users with technical expertise and guidance to design and implement modern active archive strategies that solve data growth challenges through intelligent data management.

FOREWORD

We would like to take this opportunity to welcome our readers of this 2023 Active Archive Alliance annual report and to say a special thanks to our members for their insights and contributions.

Also, let us extend a special welcome to our newest members who represent truly diverse products, services and markets. These newcomers speak to the widespread challenges facing so many organizations today when it comes to relentless data growth and how to effectively manage it.

Drawing on recent market reports and feedback from data management teams, we can safely say that the tactical measures of simply buying more storage are not sustainable. By the end of this decade, we may very well be facing new enterprise storage capacity shipments of some 15.0 ZB with an active installed base exceeding an overwhelming 45 ZB.

As an industry, how can we supply such immense quantities of storage? As consumers of enterprise storage, how can we pay for it? How can we as a society provide the necessary power and control the associated carbon footprint? How can we protect it all in this never-ending age of cybercrime?

What's needed is a modern strategy to manage the growth and volume of data, and we believe that is the solution that active archives provide. For today's strategic, data-driven organizations, data deletion is almost never desirable, while data democratization is imperative.

We hope our readers will better understand the value of intelligent data management, the many benefits of active archiving, and why this strategy is quickly gaining popularity.

Finally, we invite you to collaborate with our members to leverage their expertise and the rich and varied products and services that they offer.

Foreword by:

Betsy Doughty, Vice President of Corporate Marketing, Spectra Logic, Co-Chairperson, The Active Archive Alliance

Rich Gadomski, Head of Tape Evangelism, FUJIFILM Recording Media USA, Inc., Co-Chairperson, The Active Archive Alliance

TABLE OF CONTENTS

"All industries are being transformed by data and every CIO should now have a strategy for 100-year data retention with the highest availability and durability at the lowest sustainable cost."

— **Dr. Bhupinder Bhullar,** Swiss Vault CEO

EXECUTIVE SUMMARY

Business and IT leaders well understand the challenges of massive, double-digit data growth. More devices and applications generate more data from the edge to the public cloud. Copying and replicating data for protection, the need to keep data for longer periods, and even the fear of deleting corporate data add to storage demands. A 40% annual data volume growth rate drives the experience for many industries.[1]

Key to handling the data growth challenge, especially in the context of flat or slowly growing IT budgets, is effective data management. A working definition of data management involves the processes for gathering and storing data efficiently, securely, and cost-effectively. Without effective data management, data growth overwhelms an organization.

> ***"Data classification will play a role in helping data owners know what they have and what medium to store it on."***
>
> — **Paul Luppino,** Director, Global Digital Solutions Data Management, Iron Mountain

Effective data management brings other crucial benefits. Numerous challenges face IT organizations today, including ransomware threats, budgetary pressures, skill set shortages, and digital transformation. Intelligent, effective data management addresses these issues through cyber resiliency, reducing costs, simplifying data administration, and data accessibility features.

Beyond problems to be solved, data-driven organizations recognize data as a strategic, enterprise asset. In a future world where AI and ML workloads permeate and drive business processes and decision-making at all levels, effective data management becomes imperative. Organizations without intelligent data management processes that feed into business intelligence workloads risk being left behind by their competitors who do.

And this is where the active archive model serves today's modern and future enterprises.

Active Archiving solves data growth challenges through:

- An intelligent data management layer to place data where it belongs for cost or performance
- Adaptability to any storage architecture, media, or protocol
- Applicability across the entire data lifecycle, from data creation through archiving and eventual purging
- Security and protection features that safeguard data from threats and risks

An active archive positions organizations to cost-effectively manage their growing data and address industry pressures, while laying a foundation to profit from tomorrow's opportunities.

> ***"Storing the increasing volumes of data will continue to be one of the priority problems for many companies."***
>
> — **Thomas Thalmann,** CEO, PoINT Software & Systems GmbH

ANCIENT ARCHIVES, ACTIVE ARCHIVES

The archive began as an inevitable outcome of written communication. Archaeology has surfaced ancient archival records of contracts, trade, law, and history. These archive records were active. History records that administrators, officials, tradespeople, and scholars accessed these archives to make decisions about the present and the future.

Fast forward the transition of archive media from tablets, scrolls, books, and paper to today's digital era with flash, solid state, hard disk, optical, and tape drive media. Archiving digital information has grown into a multi-billion-dollar marketplace of products and services.

Whereas early uses of archiving in the digital age focused primarily on moving cold data to long-term storage, today's modern active archive model recognizes the value of archived data; that, like ancient times, archived data must be active and available to end users, applications, and workflows for discovery, research, and analysis.

What is much different today compared with archives of former eras is the tremendous volume of information generated in digital form. Which is why organizations need an active archive to successfully manage, store, and access their growing volumes of data.

THE HEART OF AN ACTIVE ARCHIVE – INTELLIGENT DATA MANAGEMENT SOFTWARE

At the center of an active archive resides an intelligent data management system. This software system plays the central role of automatically placing data where it belongs for cost, performance, and workload priorities. Using technologies such as metadata and global namespaces, the data management layer makes data accessible, searchable, and retrievable on whatever storage platform or media it may reside.

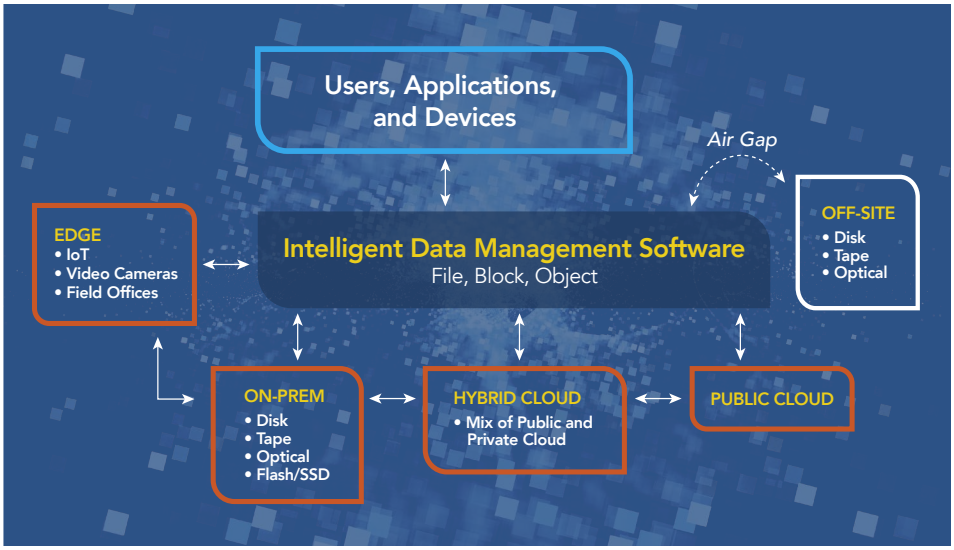
Among its many features, the intelligent data management layer adds value by:

- Automating the decisions for tiering data to long-term storage
- Automating data management processes such as:
  - Applying data protection and security policies
  - Cleansing data
  - Alerting for anomalous conditions
- Surveying and analyzing the enterprise data landscape
- Discovering data for which IT administrators are not aware
- Presenting visual representations of an organization's data through charts, graphs, and dashboards for better decision-making
- Simplifying the skill set needed to oversee and manage large, growing volumes of data

And the data management software does this work in the background without affecting performance.

THE ACTIVE ARCHIVE

Integrates Intelligent Software and Scalable Storage for the Optimal Archive Solution



IT leaders considering an active archive model will find it agnostic to the organization's storage protocols, media, or architecture. Also note, in contrast to a single vendor solution, the active archive model comprises an integrated system of products and services from different vendors using different technologies.

"File/object storage interfaces for on-premises tape archives allows users to take advantage of the reliability and economies of data tape for active archives."

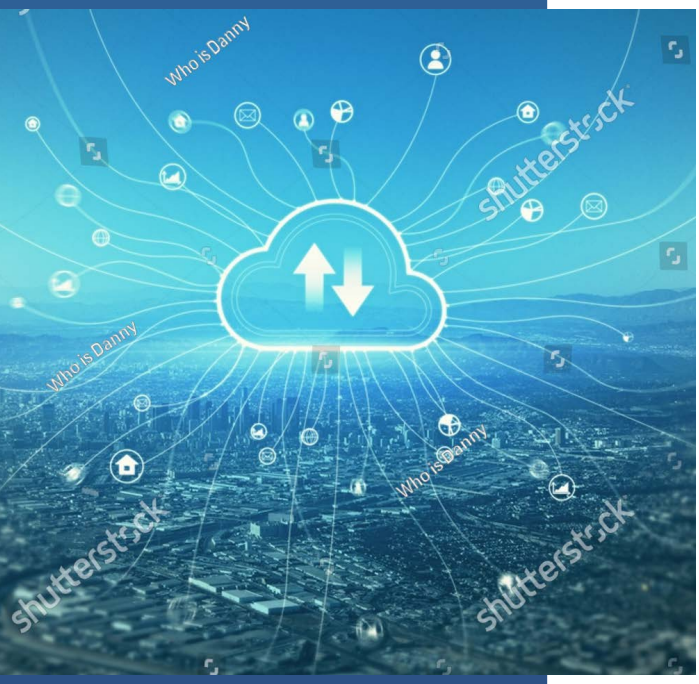
— **Phil Storey,** CEO, XenData

"Active archiving is being leveraged for capacity management to remove less used data from backups and improve system performance."

— **Betsy Doughty,** VP of Corporate Marketing, Spectra Logic

"An important benefit of SMR HDD technologies is that a single tier of storage can deliver HDD-level data access performance at $/TB levels never before realized."

— **Mark Pastor,** Director, Platform Product Management, Western Digital



THE BENEFITS OF AN ACTIVE ARCHIVE

Access

An essential active archive principle asserts that today's enterprise needs online access to legacy data. Some use cases may need fast access; in other instances, a longer retrieval time is acceptable. The organization determines its access needs and permissions for users and groups.

Access to an active archive serves the organization through the following:

- **Business intelligence:** Companies can analyze retained data for insights into trends and patterns. Monetizing data becomes the ultimate objective to transform storage costs into profit opportunities.
- **Legal requirements:** Ongoing access ensures legal teams can search and retrieve data from cold storage in response to litigation.
- **Offloads IT resources:** Configuring online access to inactive data so users can retrieve these files without IT intervention. Self-service access saves time and money.

Cost Savings

Most data growth comes through unstructured data represented by video, audio, images, presentations, email, and documents. The likelihood for users to access this data 30 days after its creation drops considerably; after 100 days, it falls below 1%.[2] To keep this growing, inactive data on primary storage becomes inefficient and costly.

Through intelligent data management software, an active archive moves inactive data to low-cost storage. For some organizations, data management software can tier older data to warm storage such as hard drives. Then when, by policy, data has aged sufficiently, files can move to more cost-effective storage such as economy disk, tape, optical, or even the cloud. Other organizations benefit by moving data immediately to archival-type storage. An example might be healthcare, where a medical image is archived immediately, but a cached copy remains on local storage for 30 days.

By moving inactive data off primary storage, the IT architecture benefits through a lean, primary storage supporting the most active data sets. The freed space on primary storage can deter the need for additional hardware purchases. Data consolidation opens cost-saving possibilities through reduced labor costs, licensing fees, and energy costs. Backup costs for primary storage also benefit as backup software has a reduced workload, needing less time and energy for its work.

An active archive saves money through data insights. The analysis provided through data management software helps IT decision-makers understand how and why their organization uses data. This understanding, in turn, shapes how administrators optimize their data for cost savings. Through trend analysis, data intelligence helps IT leaders plan and budget for storage growth.

Legal teams find an active archive saves money through cost avoidance. Most corporations must comply with regulatory requirements to keep and store data in a specific way for a specific time. An active archive helps companies comply with these data security requirements and avoid legal costs from non-compliance.

Security

The threat of a successful cyberattack worries organizations around the world. Over the last two years, ransomware has remained the number one security concern for chief information security officers.[3] A successful ransomware attack can result in data loss, business interruptions, revenue losses, fines, and legal fees. Added up, the average total expense for business recovery from a ransomware attack: $1.85 million.[4]

An active archive can supply a wide range of security features and cyber resiliency capabilities to secure and protect data from cyber threats facing today's businesses and institutions.

Examples include:

- Encryption
- Multi-factor authentication
- Access Controls Lists (ACLs)
- Role Based Access Controls (RBACs)
- Zero-trust security models

Because archival data typically remains unchanged, administrators may use WORM or view-only mode features to prevent data from being deleted or overwritten, safeguarding data's integrity, availability, and confidentiality.

> ***"New and existing regulatory aspects of data storage retention will require a different look at governance options."***
>
> —**Kel Pults,** Chief Clinical Officer & VP Government Strategy, MediQuant

As data management software moves inactive data onto active archive media, the exposure target for malware infection decreases for primary storage. Further, several media technologies, such as tape or optical, feature easy-to-deploy air-gap defenses, where IT personnel can establish a literal separation from any online path to prevent unauthorized electronic access.

"A cloud-based archive solution allows healthcare organizations to consolidate data from multiple disparate systems into a single system, supporting interoperability, data accessibility, and better patient care."

— **Ajay Kapare,** Chief Strategy and Marketing Officer, ELLKAY

Storage administrators can employ 3-2-1-1-0 best practices for their active archive data, which is also a best practice for backup storage:

- Maintain at least 3 copies of data - where the primary archive file counts as one of these copies
- Store 2 of the copies on different media (e.g., tape and HDD)
- Ensure at least 1 of the copies is stored offsite
- Store at least 1 of the copies offline
- Verify the copies have 0 errors or virus infections
- And periodically test for restoration!

> ***"An organization's data is one of its highest-value assets and the target for state-sponsored actors and cybercriminals."***
>
> — **Jason Lohrey,** CEO and Founder, Arcitecta

While organizations should depend on cybersecurity software as a first line of defense against malware, they should presume that a successful attack can occur at any time. When dealing with massive data growth, which only expands the attack surface for cyberattacks, these capabilities and practices ensure an organization's data assets remain secure, protected, and recoverable.

"Active archive solutions solve many of the pressing data storage challenges including now vitally important sustainability goals."

— **Rich Gadomski,** Tape Evangelist, FUJIFILM Recording Media, USA Inc.

"Storage teams will implement more automated data tiering to less energy-intensive storages that are more in line with data usage frequency."

— **Ferhat Kaddour,** VP Sales & Alliances, Atempo

"We all play a part in keeping this world livable. Increasingly, this is the expectation that customers bring to the table."

— **Colin Presly,** Seagate's Senior Director at the Office of the CTO

"Tape's energy efficiency and low TCO make it both a smart and green choice for active archiving."

— **Eric L. Kelly,** Chairman and CEO, Overland-Tandberg

"Enterprise-scale optical storage will help fuel sustainable long-term storage growth in the data center industry over the next decade and beyond."

— **Steve Santamaria,** CEO, Folio Photonics

ACTIVE ARCHIVES FOR EMERGING TRENDS

In addition to access, cost, and security benefits, an active archive provides technical leaders the flexibility to adapt to new industry trends and growth areas like sustainability, artificial intelligence, and edge computing. This flexibility helps enterprises thrive with possibilities for market leadership, increased revenue, and competitive advantages.

Sustainability and Active Archives

From an IT perspective, sustainability concerns itself with how corporations use IT systems to minimize their negative impact on the environment and society while maximizing their positive impact. Sustainability focus areas for IT include energy consumption, e-waste, and supply chain efficiency. Above legal requirements for sustainability, business leaders recognize the public is more likely to support companies committed to products and services that help the environment. Further, sustainability practices that save energy and reduce waste, save money.

Data centers consume a significant amount of the world's energy, with estimates as high as 3% of the global energy supply.[5] Servers and storage are of particular note for energy consumption. Each new year of data growth adds to the energy needs required to run servers, storage systems, and networking equipment for processing, storing, and transmitting data. An active archive's adaptability to an organization's priorities makes it well-suited for supporting sustainability goals.

For example, and as noted before, the intelligent data management software layer automatically places data for cost, performance, and workload priorities. By tiering inactive data off energy-intensive devices such as flash or performance HDDs, inactive data can move onto less energy-intensive technologies like (some) HDDs, tape systems, and optical storage, known for their minimal energy requirements and low cost per terabyte.

By virtualizing the underlying storage infrastructure, data management software optimizes storage resources for energy savings. Reporting and analysis from its software guide decisions for data that can be retained or purged for energy and cost savings. Analysis can help enterprises consolidate their storage resources to reduce energy consumption. Enterprise leaders can use their reports to link storage technologies to energy consumption. These analysis features contribute to reporting requirements organizations may have for sustainability activities.

And so, through an active archive, intelligent data management becomes intelligent energy management.

CLOUD REPATRIATION

There is little wonder that the public cloud services market is expected to grow to $1 trillion worldwide by 2026.[7] The value proposition to corporations is quite simple: With a credit card, scalable infrastructure is instantly available. This instant infrastructure is not without challenges:

- **Cloud cost overruns.** Through 2024, nearly 60% of companies will experience public cloud cost overruns that negatively impact their budget.[8]
- **Data sovereignty.** Cloud storage may impact data sovereignty regulations for which a corporation is legally responsible.
- **Cloud data breaches.** Nearly half of organizations have experienced a cloud-based data breach or failed audit in the past 12 months.[9]
- **Cloud latency.** Certain workloads or data transfers may have time-sensitive requirements.

Certainly, there are benefits to cloud storage and services. The issue is that cloud services do not scale cost-effectively. And it is why many corporations are repatriating some or even all of their long-term data to on-premises, private cloud, or hybrid-cloud solutions.

Active Archives for AI/ML frameworks

Artificial intelligence (AI) and machine learning (ML) workloads will permeate the workplace as an enterprise tool for operations management and decision-making at all levels. Market research indicates 35% of organizations have already invested in AI, and 44% plan to invest in AI within the next year.[6]

As effective data management improves AI, effective AI improves data management. AI expands the intelligent data management software layer beyond analysis and reporting. AI will bring value to data management and, therefore, to an active archive through:

- Tailored optimization for primary, backup, and archival storage for availability, cost, performance, and workload priorities
- Applying and enforcing data security and protection protocols
- Monitoring, detecting, and countermanding security threats
- Increasing the scope and productivity of IT employees
- Accelerating and automating complex data management processes
- Automating data recovery of critical workloads in the event of cyber attacks or other disruptions

> ***"Reducing energy consumption improves the fiscal bottom line and results in a lower CO2e contribution from infrastructure."***
>
> — **Kiyoshi Urabe,** Business Line Executive Data Retention Infrastructures Product Manager, IBM

Through an AI-driven intelligent data management software layer, AI will automate and autonomize active archives by:

- Automatically cleansing, normalizing, classifying, and making accessible long-term data and metadata for AI workloads
- Automating metadata tagging, indexing, and data cataloging inactive data
- Identifying and archiving sensitive information
- Presenting Natural Language Processing (NLP) interfaces for archive inquiries
- Efficiently allocating compute, storage, and networking resources in support of sustainable efforts

Ultimately, AI depends on well-organized data for success. Which highlights yet again why effective data management through an active archive is crucial for an AI future.

> ***"Artificial Intelligence (AI) will alter the landscape of unstructured data—forever."***
>
> — **Bruce Kornfeld,** Chief Marketing and Product Officer, StorMagic

Edge Computing and Active Archives

Edge computing moves data processing out of the data center and at or near the source where it originates. Data processing may be in a device or a small server. Fueled by IoT, 5G, and advances in small form factor server technologies, the edge computing market is expected to rise to $116.5 billion by 2030.[10]

Computing at the edge brings IT benefits through fast responses, reduced network transmissions, and reduced costs. The distributed nature of edge computing complicates data storage, where edge devices may number from hundreds to hundreds of thousands. These devices may range from small sensors to small servers. Enterprises are faced with questions about collecting, processing, storing, and managing the data generated from edge devices and applications.

> ***"We will see more organizations using* Archive from the Edge *technology."***
>
> — **Dave Thomson,** SVP Sales and Marketing, QStar Technologies

An active archive adds value to edge computing in many of the same ways as data generated within the data center and cloud. The intelligent data management layer helps IT organizations with managing edge data through:

- Automatically placing edge data where it belongs for cost, performance, and workload priorities
- Automating data management processes such as:
  - Applying data protection and security policies to edge data
  - Tiering edge data for cost and workload priorities
- Analyzing and presenting a visual representation of edge data through charts, graphs, and dashboards for better decision-making

Additionally, some edge data may be eligible to be moved to an active archive environment immediately. Edge data to be retained can be stored in a WORM or otherwise immutable format to protect it from malware. IT organizations can use their active archive to make edge data readily available when needed for analytic workloads for business insights.

VIDEO SURVEILLANCE

One of the many use cases of edge computing involves video surveillance for public security. Video surveillance installations are an ideal candidate for collecting, retaining, and accessing recorded video data volumes.

With estimates of over a billion devices worldwide, a large international airport may generate hundreds of terabytes of video surveillance data each day. Even an organization with a hundred cameras can generate tens of terabytes daily.[11]

An active archive serves as a practical video surveillance storage solution for organizations requiring long-term video retention or having large numbers of surveillance cameras.

Management: An active archive's intelligent data management policy can tier video files onto HDD and then to an active archive storage tier.

Costs: Tape technologies bring cost-effective, scalable solutions for active archive video surveillance footage.

Security: An active archive, through its data management software, can automatically apply security policies to ensure regulatory compliance and protection.

Access: With integrated tiering into video management software (VMS) systems, the ability to search and playback all recorded video from either tier of video storage becomes straightforward for the video operator.

Active archive, multi-tier, video storage solutions bring cost-effective implementations for video surveillance use cases. These solutions significantly reduce the initial cost of required hardware, plus the overall total cost of ownership benefits from power cost savings.

THE ACTIVE ARCHIVE IMPERATIVE

The Information Age floods the modern enterprise with data. This exponential demand trend for enterprise storage capacity will only increase in the years ahead. This massive growth is why the urgency and priority of effective data management from the edge to the cloud are imperative.

At a minimum, the infrastructure of the new, upcoming world of artificial intelligence will be laid on a foundation of well-organized data stores and workflows. Even today, many AI projects are shelved because the data was never collected. Further, a lack of skilled data management will impact IT costs, data security, cyber resiliency, legal compliance, customer experiences, decision-making, and brand reputation, among others.

Proficient data management advances a data-driven culture. That is a culture where organizations treat data as a strategic asset for big decisions and everyday actions. And like AI, data management links foundationally with data-driven values. Still an aspiration for most companies, as only a quarter of organizations report they have established a data-driven organization.[12]

Ultimately, effective data management is a necessary part of core competencies organizations must achieve for IT modernization and digital transformation.

"The sheer magnitude of data in the next era of digital business requires organizations to rethink the scale, cost, durability, and intelligence underlying their data management strategies."

— **Eric Bassier,** Sr. Director, Product Marketing, Quantum

"Businesses are discovering that their archived content has new analytical or regulatory value."

— **Derek Dicker,** CEO, Nyriad

And this is where the active archive model serves today's modern and future enterprises.

The active archive solution answers the need for effective data management while providing a cost-effective, scalable solution addressing data growth challenges. Beyond a solution for data growth and management, an active archive serves the organization and its stakeholders by advancing the broader initiatives for digital transformation.

"Legacy data is the hardest to remediate and purge. With active archives legacy data problems are easier to resolve."

— **Brendan Sullivan,** CEO, Sullivan Strickler

For more information on the ACTIVE ARCHIVE ALLIANCE or to learn more about an industry-specific active archive solution to meet your needs, visit:

activearchive.com

THE ACTIVE ARCHIVE ALLIANCE

The Active Archive Alliance serves as a vendor-neutral, trusted source for providing end users with technical expertise and guidance to design and implement modern active archive strategies that solve data growth challenges through intelligent data management. Active archives enable reliable, online, and cost-effective access to data throughout its life and are compatible with flash, disk, tape, optical, or cloud as well as file, block or object storage systems. They help move data to the appropriate storage tiers to minimize cost while maintaining ease of user accessibility.

Active Archive Alliance Members & Sponsors

FUJIFILM
SPECTRA
MediQuant
IRON MOUNTAIN
OVERLAND TANDBERG
ATEMPO preserving data ecosystems
POINT software & systems
Western Digital
IBM
Quantum
Folio Photonics
QSTAR TECHNOLOGIES SECURING THE FUTURE OF YOUR DATA
ELLKAY
XenData
StorMagic
SWISS VAULT
Sullivan | Strickler
NYRIAD
SEAGATE
ARCITECTA

Sources:

1 https://www.statista.com/statistics/1186304/total-enterprise-data-volume-location
2 https://www.enterprisestorageforum.com/cloud/guide-to-google-cloud-storage
3 https://www.zdnet.com/article/ransomware-is-still-the-biggest-security-worry-for-business-but-its-not-the-only-headache
4 https://www.sophos.com/en-us/medialibrary/pdfs/whitepaper/sophos-state-of-ransomware-2021-wp.pdf
5 https://www.bu.edu/hic/2022/10/12/ask-the-experts-how-are-data-centers-impacted-by-climate-change
6 https://www.infotech.com/research/ss/tech-trends-2023
7 https://www.prnewswire.com/news-releases/forrester-public-cloud-is-poised-to-surpass-1-trillion-by-2026--but-not-without-enduring-several-global-challenges-301701469.html
8 https://www.informationweek.com/cloud/10-cloud-strategies-to-avoid-cost-overruns
9 https://www.scmagazine.com/news/cloud-security/nearly-half-of-businesses-had-a-cloud-based-data-breach-or-failed-audit
10 https://www.globenewswire.com/en/news-release/2022/07/07/2475942/0/en/Edge-Computing-Market-Size-to-Hit-at-USD-116-5-Billion-by-2030.html
11 https://cozaint.com/askalice
12 https://hbr.org/2022/02/why-becoming-a-data-driven-organization-is-so-hard