SEAGATE

eBook

A Better Way to Manage Geosciences Data at the Edge

The right mass data transfer solution can expedite data processing for the Geosciences industry.

Table of Contents

Goal: Make Data Useful Faster

Today's geoscience companies rely heavily on data to inform every segment of their workflow, whether that information is seismic, oceanographic, meteorological, or structural. This massive volume of real-time data comes from a plethora of devices in the field like drones and sensors—often operating at the extreme edge—and is critical to exploration, worker safety, asset maintenance, and operational efficiency.

Raw data from the field is processed and analyzed in many different ways. In some cases, initial analysis occurs immediately upon ingest, enabled by enterprise-grade, mass storage at the site. In other instances, key data travels to an edge data center for more sophisticated processing and timely use. Sometimes subsets of data are transported to different locations for use by unrelated teams solving unique problems. Often the bulk of the data is moved to a cloud content repository where it waits for the initial processing of smaller data sets to illuminate patterns that inform later-stage analysis.

In geosciences, data is processed at various points to deliver actionable intelligence, so an efficient and secure solution is essential, no matter where data is captured or where it is going.

Let's look further at the data challenges in geosciences, and the mass data transfer technologies available for this industry.

Unique Data Collection and Transfer Problems

Geosciences is a data-hungry industry. Dataset volume has skyrocketed, and data growth continues to climb sharply. Massive amounts of raw data are collected and stored at the extreme edge—often in remote locations and harsh environments. Processing and analysis can happen at various points along the data lifecycle—in the field (i.e., on seagoing vessels or hard-to-reach geographies), at an edge facility, or in a central data center. Handling, transferring, and transporting this data properly is critical to its integrity and ultimate value. Geoscience companies need a data transfer solution that offers:

- Fast data ingest
- Simple connectivity from capture device to ingest
- Physically safe and viable data
- Secure data
- Compatibility with existing enterprise-level data center and cloud systems

Data's Value in Geosciences

In geosciences data from the field is the most valuable asset. Every byte matters because any data left out or lost equals a missed opportunity. The more data collected, the more accurate the insights from the analytics team. All varieties of capturable unstructured data from multiple data streams need to be preserved as well. In geosciences, data transfer must be complete, with no exceptions.

Data is a tool to derive insights, and because data loses value every day and can quickly become outdated, it's crucial for geoscience companies to shorten the timespan between data capture, ingest, transfer, processing, and analysis in order to deliver actionable insights to customers faster. When this is done, it can be a game changer for operating efficiency and customer satisfaction in a highly competitive industry.

Unique Challenges of Data Storage and Transfer at the Extreme Edge

Remote field sites in far-flung locations at the extreme edge (e.g., oceans, deserts, mountains) are another factor. It can be difficult to move large data sets quickly at these sites because of limited network access and bandwidth. In addition, harsh conditions are often the norm so physical equipment concerns are significant in these environments. If data is loaded onto multiple hard drives, it can be challenging to keep track of all the devices during transit. When lower-capacity hard drives are used it can also mean that initial processing can't be done at the site. Transporting storage devices in heat and over rough terrain or long distances creates legitimate worry about data integrity and damage.

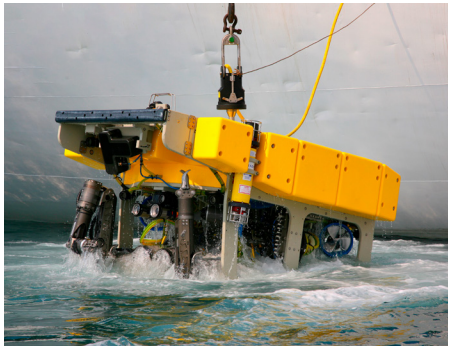
Let's look at a few real-world use cases.

Use case: 10,000-mile transcontinental natural gas pipeline inspection

- **Purpose:** Monitoring for cracks and fractures
- **Device:** Fleet of PIGs (pipeline inspection gauges) in constant motion within the pipeline collecting data from electromagnetic sensors
- **Frequency:** Data downloaded from PIGs every two days
- **Raw field data volume:** 41TB/trip and 2.1PB/annually

Use case: Subsurface exploration on land

- **Purpose:** Use distributed acoustic sensing (DAS) to get subsurface data
- **Device:** Fiber optic cables gathering data on subsurface conditions in underground geologic structures like wells
- **Frequency:** Weeks and months in the field
- **Raw field data volume:** 20-30TB/day and up to 1PB for a single job



Use case: Seabed data

- **Purpose:** Seismic survey for oil/gas exploration, general intelligence for wind farm development
- **Device:** Numerous collection points at the edge (drone swarms or nodes on a rope from 25 vessels)
- **Frequency:** Two-month-long voyages
- **Raw field data volume:** 1-3PB/voyage

Use case: Autonomous surface vessel

- **Purpose:** Climate or ocean monitoring
- **Device:** Uncrewed ocean-going drone
- **Frequency:** Six–to-12-month deployments
- **Raw field data volume:** Up to 1PB for a single project

Existing Legacy Data Transfer Models

Current data transfer models have many shortcomings. Because data sets are so large now, typical bandwidth limitations at the edge where geosciences data is collected and ingested means the transfer time is painfully slow, sometimes measured in days. Standard equipment is bulky, unwieldy, or fragile. Considering the environment it must function in and travel through on its way to a final destination, neither characteristic is acceptable. Sometimes the equipment, such as a fleet of smaller hard drives, lacks sufficient capacity for the massive data sets generated in the industry.

Security is an issue, both in terms of worries about tampering as well as conforming to data security regulations and norms. Most current solutions offer narrow options for vendors and platforms, where the vendor ecosystem is dictated—by either a specific vendor or limited choices—and the platform is restrained. Also, a major drawback is that many solutions require a significant upfront purchase which means a high initial capital expense often followed by pricing and contract inflexibility.

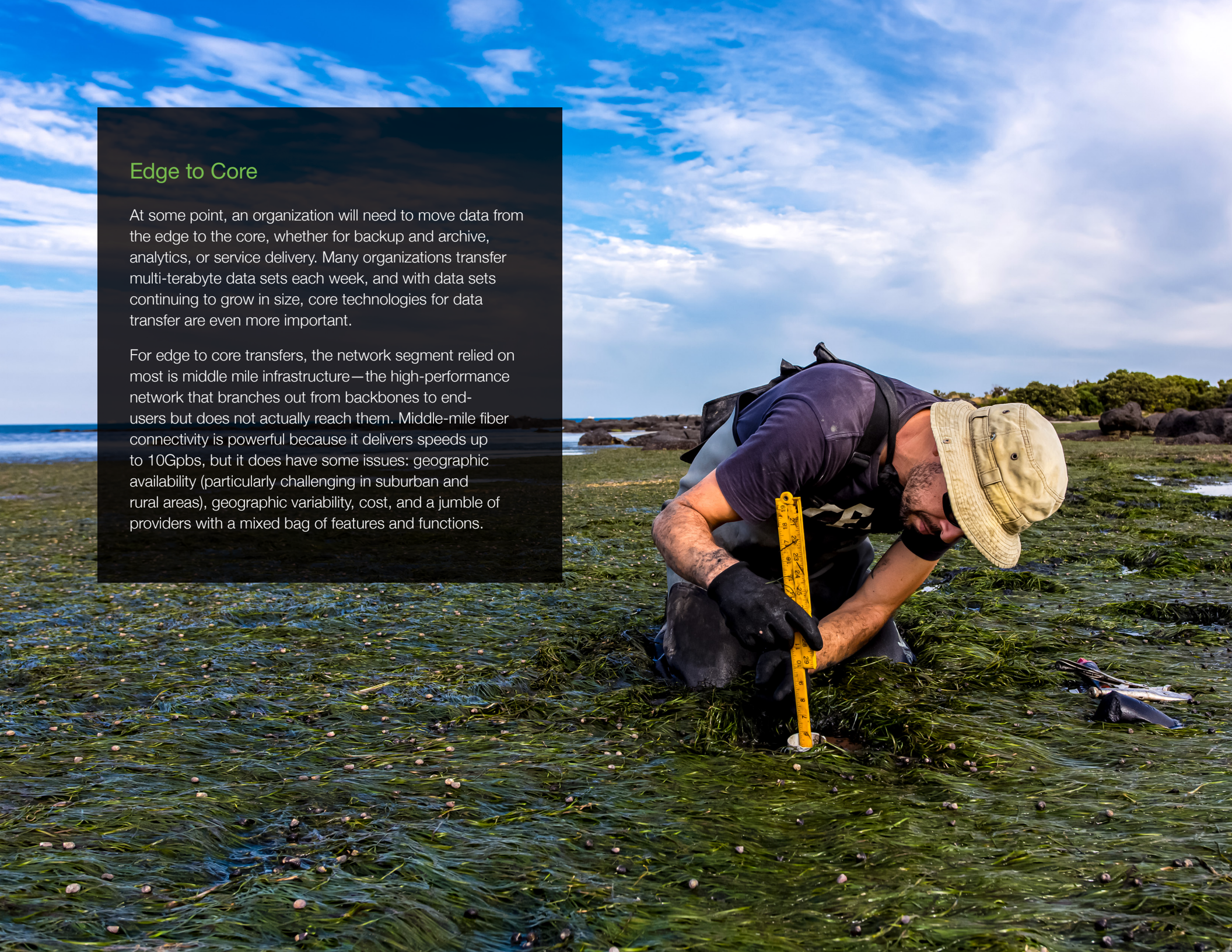
Here are some existing legacy options data managers and IT architects in geosciences have when capturing, processing, and moving data between endpoints, edge, core and cloud:

Endpoint to Edge

The endpoints—the capture devices in geosciences—are the source of all data. They typically connect to edge infrastructure through "last mile connectivity," whether wired or wireless. Because of the massive size of data sets in geosciences and the harsh environment where these endpoints are required to function, many data movement technologies fall short. For instance:

- Endpoint to edge connections that rely on wireless networks often have issues with speed, consistency, reliability, geographic availability, and cost.
- Endpoint to edge connections using low-Earth satellites, a resurging option first tested in the 1990s, are hamstrung by delays in receiving equipment, consistency (particularly when cloud cover is present), speed, hemisphere availability, and cost.
- Endpoint to edge connections with last-mile wired network infrastructure (e.g., DSL, cable, coax, or fiber optics) deliver consistent, enterprise-grade performance but also suffer from limitations such as access (differs widely by state and county), speed (degrades as soon as you exit a major metropolitan area), and cost.

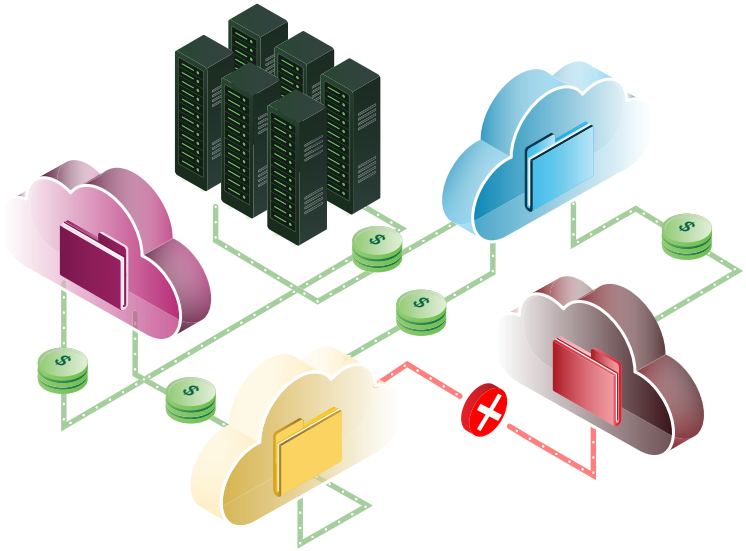
In the real world, most organizations use a combination of solutions to manage their endpoint-to-edge data transfer. But with the shift toward distributed computing and the trajectory of data growth, pressure continues to increase on existing solutions. Many companies tired of seeking workarounds hope for an alternate technology that can move multi-terabyte data sets from anywhere, as often (or infrequently) as needed, reliably, with enterprise-grade performance and technical simplicity to eliminate obstacles in the field.



Edge to Core

At some point, an organization will need to move data from the edge to the core, whether for backup and archive, analytics, or service delivery. Many organizations transfer multi-terabyte data sets each week, and with data sets continuing to grow in size, core technologies for data transfer are even more important.

For edge to core transfers, the network segment relied on most is middle mile infrastructure—the high-performance network that branches out from backbones to end-users but does not actually reach them. Middle-mile fiber connectivity is powerful because it delivers speeds up to 10Gpbs, but it does have some issues: geographic availability (particularly challenging in suburban and rural areas), geographic variability, cost, and a jumble of providers with a mixed bag of features and functions.



Edge and Core to Cloud

Because the cloud is the ultimate destination of data, most enterprises need to transfer large data sets into the cloud, whether public, private, or combination environments. Cloud providers offer enterprise-grade capabilities—massive bandwidth, dedicated connections, and sophisticated mass data transfer functionality. But they come with two major drawbacks: complicated price models and confusing cross-cloud data movement cost matrices. Moving data into a proprietary cloud is simple and cheap, but moving data around or away once it's there (cross-cloud or from one storage type to another) quickly becomes an expensive, tangled mess.

For specific detailed scenarios of current technology challenges with endpoint to edge, edge to core, and edge + core to cloud, please see Enterprise Data Transfer Playbook.

Exploring Physical Mass Data Transfer

Because of edge data transfer challenges like bandwidth limitations or unavailability, dropped connections, or unacceptably high solution costs, many organizations are turning to an enterprise-grade physical data transfer solution to mitigate these issues. This solution uses physical mass storage devices capable of transporting bulk data via package carriers rather than moving data via radio waves, light, or electrons down wires.

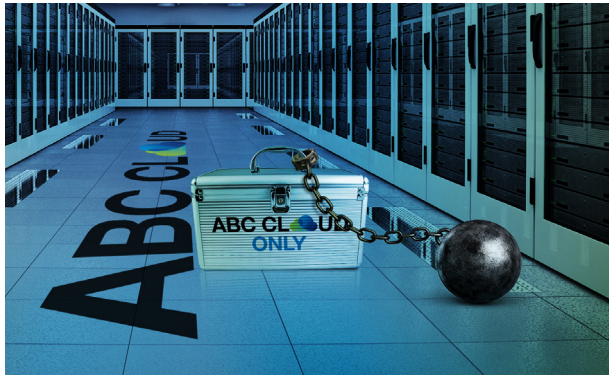
Mass data transfer platforms typically include a high-capacity drive or drives, an interconnect, and software to facilitate data transfer. The devices are shipped to a data collection or ingest location (often in the field or at an edge data center), an administrator transfers data to them, the devices are physically moved (driven, flown, or shipped) to a destination (often a core data center or cloud provider), and then data is removed from the mass storage device that can be shipped back for reuse.

Key criteria to consider in a mass data transfer platform includes:

- Capacity
- Time-to-ingest
- Time-to-access
- Frequency
- Security
- Ruggedness
- Open standard technologies
- Cost

The following explores a few examples of physical mass data transfer tools and their pros/cons

External drives are consumer-grade products with basic capacity and security measures. Performance can be unpredictable, the devices are relatively fragile for harsh environments, and since they're an outright purchase, there is an upfront cost involved.



Cloud vendor data transfer devices are enterprise-grade, so they offer excellent capacity, security, ruggedness, and portability. However, because they're offered by a single cloud vendor, they typically involve vendor lock-in (no support for "other" clouds), they don't support bidirectional data movement (they only support movement from device to cloud, not the reverse), and their cost structures are extremely complicated.

Enterprise data transfer as a service using physical shuttles is a relatively new offering capitalizing on the benefits of an "as a service" or subscription model. This solution offers enterprise-grade capacity, performance, and security paired with rugged devices, simple in-field transfer, an open standard (compatible) infrastructure, and the attractive "as a service" pricing model that is customizable and scalable according to need.

Existing Technologies in Mass Data Transfer

Technology	Transfer Speed	Strengths	Weaknesses
Wireless	External drives	• Availability • Many providers • Security • Cost	• Speed • Consistency • Reliability • Access
Wired (Last Mile or Mid-Mile)	25Mbps - 800Gbps	• Mature technology • Availability • Reliability • Many providers • Security • Cost	• Access • Performance • Cost
Cloud	Up to 100Gbps	• High performance	• Vendor limited • Complex costs
Cloud vendor device	Up to 3.2Gbps	• Enterprise-Grade • Mature • Rugged	• Vendor lock-in • No bidirectional transfer • Complex costs • Needs on-premise admin
External drive	Up to 2.32Gbps	• Availability • Open standards • Cost	• Not rugged • Security • Needs on-premise admin
Data shuttle	Up to 22.4Gbps	• Flexible • Scalable • Open standards • Rugged	• Needs on-premise admin

Why Lyve Mobile

As a leading provider of enterprise storage for more than 40 years, Seagate® now offers a uniquely nimble way for businesses to aggregate, store, move, and activate their data enterprise-grade solution as a service. Seagate's Lyve™ Mobile data transfer as a service is a modern mass data transfer solution with several unique features. First and foremost, Lyve Mobile is a simple and secure solution. Because it's rugged and can operate at the edge—even the extreme edge—it functions without infrastructure or bandwidth limitations. Lyve Mobile is purposefully designed to be vendor-agnostic so you can use it with any cloud service. It interfaces with any existing data center storage and service systems, so it truly delivers a frictionless experience. Seagate's subscription-based pricing structure aligns with the modern IT architecture model of decentralizing data infrastructure and data management to provide greater customer control, data freedom, and transparent pricing. The "as a service" model offers these benefits:

- Scalable and modular on-demand, pay-as-you-go storage means companies only pay for what they need
- Month-to-month commitments don't require an initial capital outlay or cumbersome multi-year contracts
- Lower total cost of ownership, more efficient storage costs, and a predictable monthly storage budget
- Easier physical device management and faster equipment refreshes and upgrades

LYVE™ Mobile

Lyve Mobile Benefits Across the Data and Business Teams

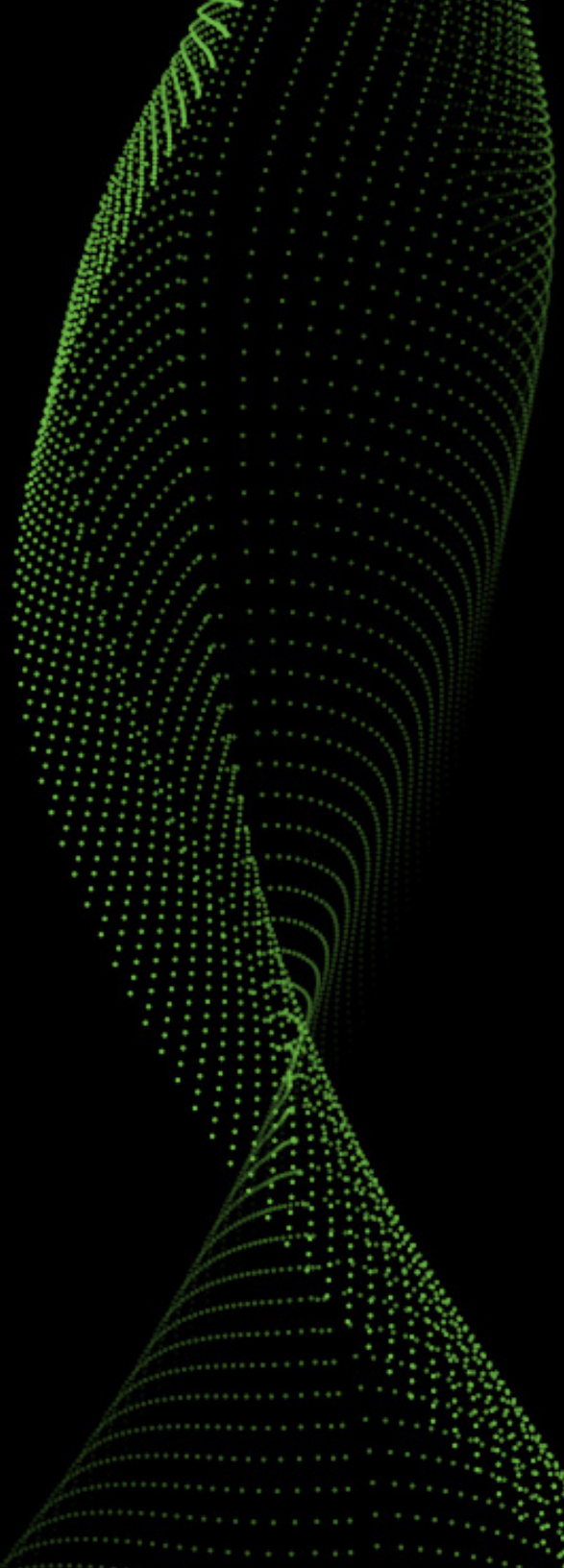
- Data managers appreciate Lyve Mobile because it easily and securely moves data from place to place so it's available to the right team for the right process.
- Field engineers appreciate how simple it is to move data from their capture devices into a Lyve Mobile system and how seamless it is to direct the data to where it needs to go after ingest.
- The data center infrastructure team appreciates the simplicity of connecting the Lyve Mobile units with their existing system and the ease of directing the data to where it needs to go after it comes off a Lyve Mobile device.
- Data analysts and data end-users benefit by being able to more rapidly access unstructured or complete data and more easily define and direct actions for the data once it enters the core's system, so it can be used quickly within their AI systems for actionable intelligence.
- Business managers see increased productivity in the field and in data processing, both of which allow nimbler creation of usable data products for the customer. Overall, a company can create higher customer satisfaction and serve more customers.

Key Takeaways

The geoscience datasphere is exploding. And in far-flung locations where bandwidth and harsh conditions are a problem, physical data transfer as a service offers a consistent, affordable solution. The future for geoscience is high-capacity edge storage infrastructure in a compact and rugged package that has a scalable and affordable pricing structure and also offers a frictionless physical transfer to any cloud service.

Seagate's Lyve Mobile modern edge storage and data-transfer-as-a-service solution allows users to consolidate and store data in any environment, process at the edge or near-edge, if needed, and quickly and safely move data to one or more ultimate destinations for further processing and analysis. At critical touchpoints in the data's journey, Lyve Mobile helps geosciences maximize data's value and ability to inform decision-making and drive successful business outcomes.

To learn more or talk with an expert at Seagate please visit: seagate.com/products/data-transport.



Resources

[1] https://www.seagate.com/solutions/edge/edge-storage/

[2] https://www.seagate.com/products/data-transport/

[3] https://www.seagate.com/resources/lyve-mobile-time-savings-calculator/

[4] https://www.seagate.com/resources/terradepth-case-study/

Ready to Learn More?

Visit us at
www.seagate.com