



Data storage costs out of control?

Learn the top 3 ways MongoDB Atlas on AWS keeps
unstructured data storage in check

Table of contents

- 3 The untold story of unstructured data
- 4 The beauty of an elegant document database and MongoDB Atlas
- 5 3 ways to reduce storage costs and complexity with MongoDB Atlas on AWS
- 8 Nesto Software cuts storage costs by 60% with MongoDB Atlas Online Archive
- 9 Get started with a free trial

The untold story of unstructured data

Unstructured data forms about 80% of big data.¹

But what constitutes unstructured data? And is it really unstructured?

Although unstructured data has an internally predefined structure, it does not follow a fixed data model. Unstructured data generally won't fit into a traditional database or data table. For example, many companies collect social media text, such as comments and feedback. These are unstructured data, whereas aspects like number of friends and followers are structured. For those that save email content, the body copy of the email is unstructured, but the "to" and "subject" fields are structured.

While many businesses grow their collection of unstructured data, they also face challenges in storing unstructured big data because of its high volume, variety, and velocity.

This ebook looks at how document databases are designed to store unstructured data, as well how MongoDB Atlas and Amazon Web Services (AWS) can help you solve unstructured data storage challenges.

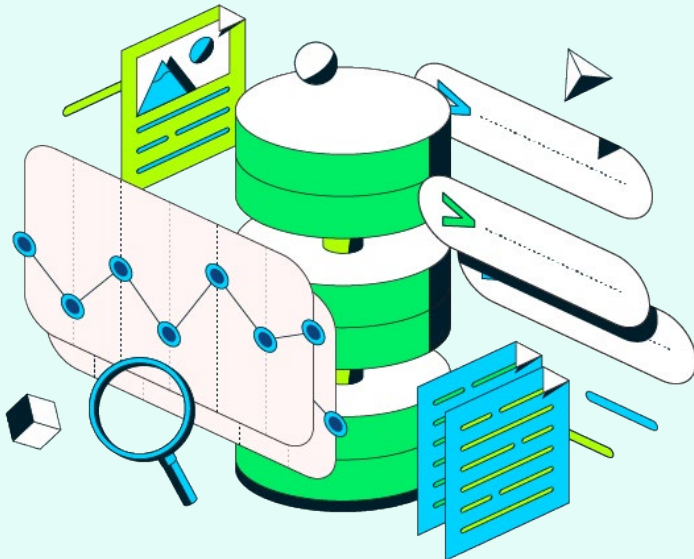
Pain points with unstructured data storage

No fixed format: Suppose you're storing details about all your employees. Each one has characteristics that others might not have, and you won't necessarily require all the fields for all employees.

Scaling issues: As the amount of unstructured data keeps increasing, traditional storage systems may not scale out. Plus, adding more resources to the system increases costs.

Complex data retrieval and querying: If you dump all your big data into a storage system, you'll need an efficient way to find, update, or even delete it, even with indexing.

¹ VentureBeat "Report: 80% of global datasphere will be unstructured by 2025" May 5, 2022



The beauty of an elegant document database and MongoDB Atlas

Unstructured data storage is complex because of the many formats it can take, as well as its high volumes. Databases are the simplest way to store data, and document databases have emerged as a convenient way of storing unstructured big data. They are flexible, scalable, highly available, secure, and help to minimize challenges with the unstructured data storage.

Document databases deliver:



Flexibility: The data model can accommodate new fields and data types with minimum impact on existing schema or data, thus requiring no downtime.



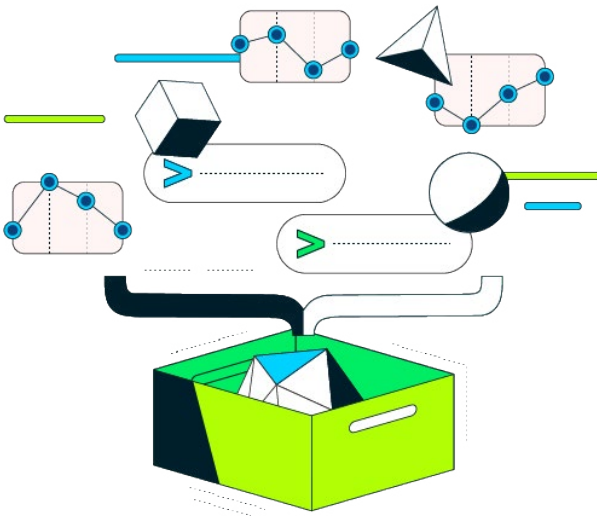
Easy access to archived data: Data archiving prevents data loss, and reduces the cost of primary storage. Data that is old but still required should be stored in such a way that it's easy to retrieve and doesn't increase overall storage cost.



Scalability: The storage system is horizontally and vertically scalable at all times without any data loss.

Ultimately, document databases make data management more efficient and cost-effective.

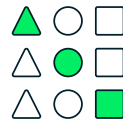
MongoDB is a document-based database, with a database-as-a-service offering called MongoDB Atlas. With Atlas, you get all the benefits of a unified database in a cloud environment, without installing MongoDB. Providing an integrated set of database and data services and a unified developer experience, MongoDB Atlas on AWS lets you build applications that are highly available, performant at global scale, and compliant with the most demanding security and privacy standards.



3 ways to reduce storage costs and complexity with MongoDB Atlas on AWS

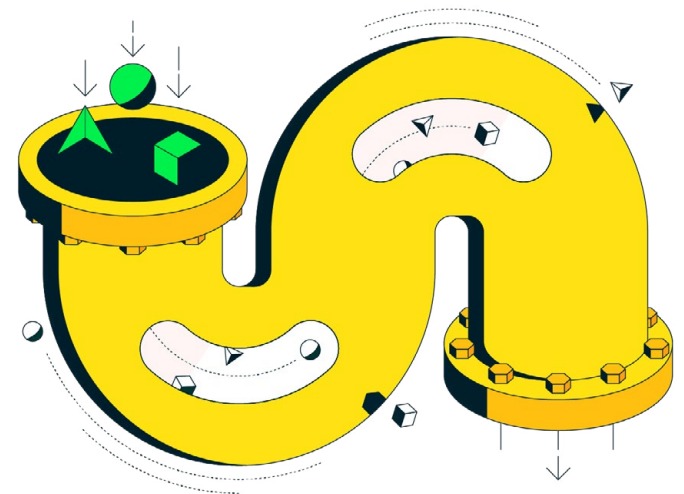
With the proliferation of different, constantly changing applications, it's challenging to innovate quickly and have a strategy in place to prevent ballooning storage costs. Legacy archival and analytic tools often restrict developers and diminish the benefits of modern, rich data sources. Developers and data engineers are also asked to apply strict schemas to data when migrating it into relational analytic tools—adding more items to their task lists. This reduces agility and increases costs, especially if developers bring in separate tooling or write their own.

With MongoDB Atlas on AWS, you can run a fully managed document database to unlock the value of your unstructured data. And because MongoDB is committed to making it easy for developers to scale and transform their data management solutions, MongoDB Atlas has several options to help you reduce storage costs and complexity.

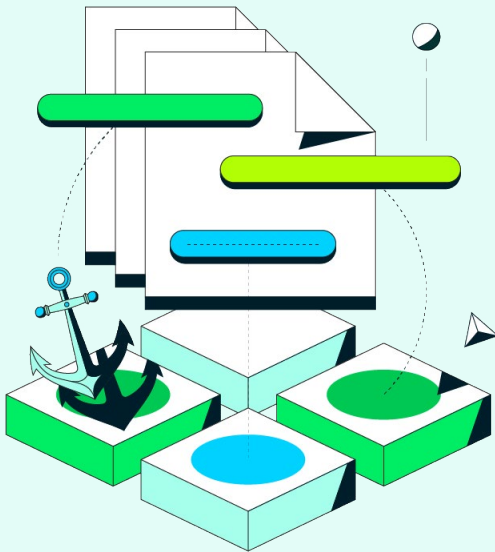


MongoDB tools like `mongodump` and `mongoexport`

Historically, customers have used tools like [mongodump](#) and [mongoexport](#) to solve data growth challenges. These are effective solutions for dropping data into cloud object storage like [Amazon Simple Storage Service](#) (Amazon S3) for retrieval in the future, or for consumption by another tool.



3 ways to reduce storage costs and complexity with MongoDB Atlas on AWS



Automated data tiering with MongoDB Atlas Online Archive

While `mongodump` and `mongoexport` allow you to get data out of your MongoDB Atlas cluster, they don't offer a fully managed and automated solution. In order to automatically export data, you would need to set up your own infrastructure, which then requires care and feeding.

With [MongoDB Atlas Online Archive](#) you can take advantage of a fully managed solution for data tiering right in MongoDB Atlas. You can create rules to automatically archive aged data to a fully managed object storage, while preserving the ability to query all your data through a single endpoint.

3 ways to reduce storage costs and complexity with MongoDB Atlas on AWS



MongoDB Atlas Data Lake and Amazon S3

Another effective approach for easily offloading aged data is using [MongoDB Atlas Data Lake](#). This is a serverless, scalable query engine that makes it easy to work with data across MongoDB Atlas and Amazon S3.

MongoDB Atlas Data Lake provides a single platform for your MongoDB Atlas clusters and allows you to:

- Organize and query across multiple Atlas and Amazon S3 clusters
- Structure the data stored in a data lake
- Convert MongoDB data into Parquet, CSV, or other formats
- Natively query, transform, and move data across Amazon S3 and MongoDB Atlas clusters

In addition to providing powerful analytic capabilities, MongoDB Atlas Data Lake is an alternative to MongoDB Atlas Online Archive for data tiering. Instead of using fully managed storage, you can archive data to your own Amazon S3 buckets. The archival process through Data Lake can be set up and automated using scheduled triggers, and data can even be written directly into Parquet format on Amazon S3 for improved future query performance.



Nesto Software cuts storage costs by 60% with MongoDB Atlas Online Archive

“It was simple to deploy Online Archive and create a low-cost cold storage environment for data, which automates archiving and tiering from live clusters into fully managed storage. It’s an affordable, reliable, and scalable way to manage our growing data needs.”

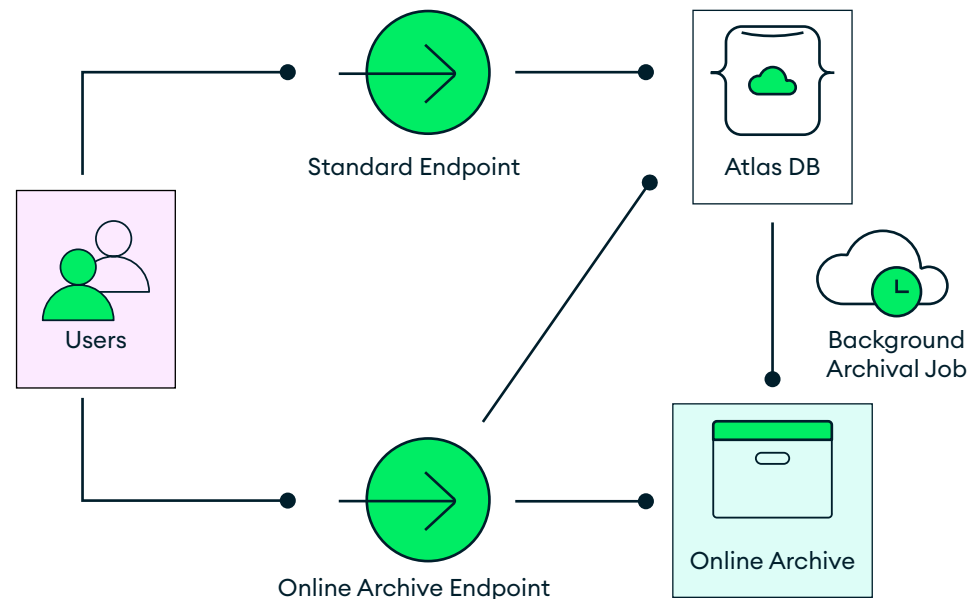
– Martin Löper, Solutions Architect, Nesto

[Nesto Software](#) is a German-based startup that powers a hospitality management platform. Part of its core functionality is forecasting staffing requirements using artificial intelligence (AI), then automating common tasks like time management and tracking employee records.

However, as the company scaled, it sought a data tiering solution that would allow it to comply with local data protection regulations, namely General Data Protection Regulation (GDPR). The team at Nesto had initially built a read auditing solution in MongoDB Atlas to comply with GDPR, where each document mapped to a read audit log. Within a single year, the audit collection had reached over 50 million documents, making up nearly 80 percent of all storage and backup costs.

To reduce costs, the team decided to set up Atlas Online Archive to offload data to more cost-effective storage, while still retaining the ability to query their data—which is a GDPR requirement. Today, Nesto has an archiving rule on any data older than 14 days, translating to approximately six million documents archived per month.

With Atlas Online Archive, Nesto saved 60 percent in data storage costs and 70 percent in cloud backup costs. This represents a reduction of 35 percent on overall database spend.



Get started with a free trial

There are many ways to manage your data footprint on MongoDB Atlas. The three options outlined in this ebook are all designed to keep you as productive as possible. Use MongoDB Atlas Online Archive to automatically move data to fully managed, query-able storage, or leverage the power of MongoDB Atlas Data Lake to archive data to your own Amazon S3 buckets.

With MongoDB Atlas on AWS, you can simplify how you work with all your cloud data. Spend less time managing infrastructure and more time accelerating development.

[Sign up for a free trial of MongoDB Atlas in AWS Marketplace](#)