

# MongoDB: A Developer Data Platform

Google slide deck available <u>here</u>

This work is licensed under the <u>Creative Commons</u>

<u>Attribution-NonCommercial-ShareAlike 3.0 Unported License</u>

(CC BY-NC-SA 3.0)

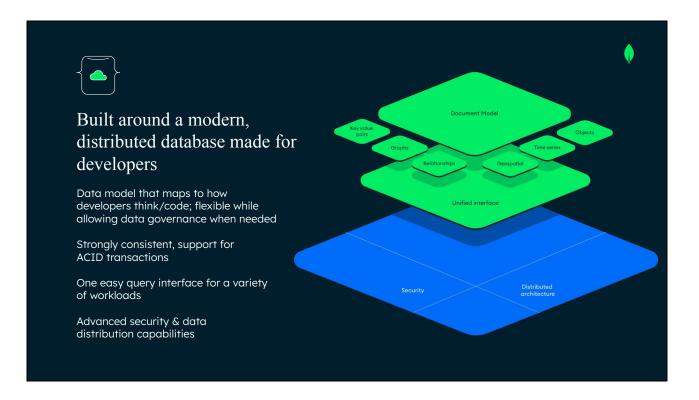


As a complete ecosystem not just a database, it supports:

Storing the data

Accessing the data

Processing the data



First, It has to be centered around a modern, distributed database built for *developers* that scales to meet tomorrow's requirements

- The platform requires a fundamentally better way to structure data for an application— the document data model is the answer. It naturally aligns with how developers think and code & modern development practices.
- It's also inherently flexible, allowing you to easily model and remodel your data for a wide range of use cases, while still enabling you to apply governance when needed.
- Features such as strong consistency & ACID transactions need to be supported. Otherwise, companies are forever chained to their relational databases.
- The Platform has to be simple to use, meaning quick to get started, next to no learning curve, with a unified way to work with many different types of workloads.
- Finally, the platform needs the advanced data distribution capabilities unlimited horizontal scale plus ability to easily distribute data to satisfy data sovereignty & ensure fast performance on a global scale



# Address a range of application use cases w/o adding complexity

Able to support full-text search functionality for delivering a fast and relevant user experience

Able to support data on mobile devices/at the edge w/o having to manually sync data

Able to deliver real-time analytics on live data w/o having to move data back & forth





# Run anywhere—start on-premises, migrate to the cloud, multi-cloud

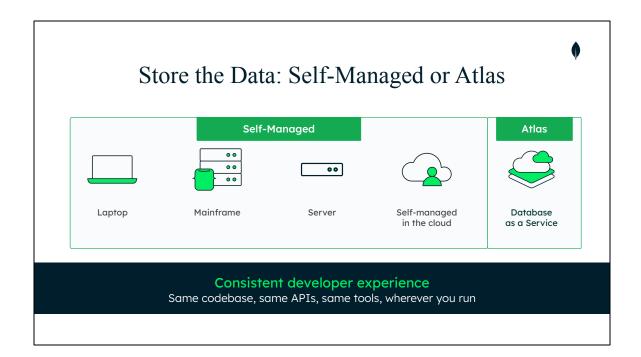
The flexibility to deploy in over 80 regions across 3 cloud providers

Database that extends across multiple clouds; get the best from each provider with no lock-in

Seamlessly migrate from on-premises to cloud w/o code rewrites



# Storing the Data



Here are the two approaches to storing your data, to clarify in either case you will need a database installation. This can either be self-managed or run in MongoDB's service, Atlas.

Self-managed means you can have the data on your laptop as part of your development environment, equally you can host it in your own server, mainframe, or yourself in a public cloud provider. The Ops Manager tool is recommended to manage, monitor, and backup these environments.

Atlas or MongoDB's "database as a service", is where MongoDB will manage and automate your deployment in one or many public cloud providers. These providers are currently Amazon's AWS, Microsoft's Azure, or Google's Cloud Platform. MongoDB and it's teams will manage and maintain your databases, in addition all of the same tooling and features available in Ops Manager are available in Atlas.





### **Atlas Cloud Database**

at the core, supporting the latest versions of MongoDB

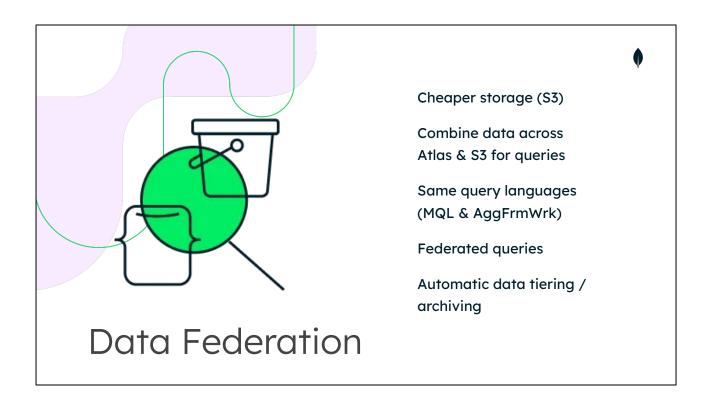
Looking into the first aspect of how MongoDB is a data platform, specifically how the data is stored. At the core, Atlas runs the latest version of MongoDB. This ensure that it can avail any new features and the upgrades of versions are handled automatically by Atlas itself.

Atlas isn't just a database alone, it is a key element of the platform so it supports and runs other tooling such as Atlas Search. This utilises Apache's Lucene to provide fine grained full-text search queries using MongoDB query language (MQL).

Atlas also links with Atlas Data Lake which provides storage on S3 and can be configured with auto-archiving to move older data, or data which is no longer relevant to your operational workload, to that cheaper storage. It can also still be queried using MQL.

Let's look at Atlas Data Lake in a little more depth.





Data Lakes are increasingly more common, it is essentially a centralized storage point for your data, whether structured or unstructured.

In the case of Atlas Dake Lake, that centralized storage is on Amazon Web Services (AWS) simple storage service (S3).

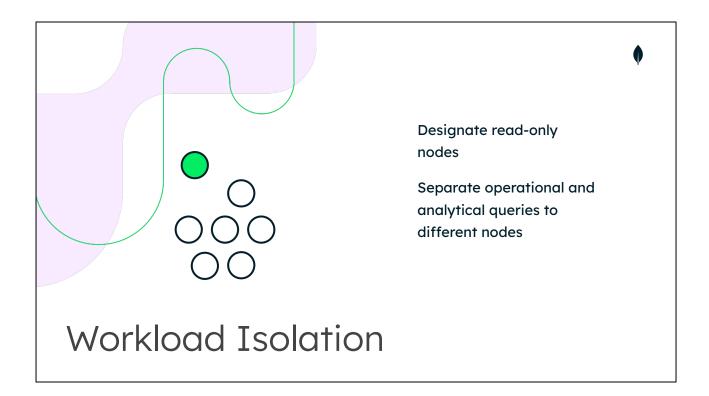
You can query across data and combine the results from S3 and from Atlas databases.

It supports a variety of file formats including JSON, BSON, CSV, TSV, Avro, ORC, and Parquet. These can be queried using the same query language used with any MongoDB database, the MongoDB Query Language (MQL) and equally you can use the Aggregation Framework. The queries can be performed using the mongo shell, MongoDB Compass, or any MongoDB driver. These queries do not trigger any data movement or transformation of the data, the queries are performed against S3.

Federated queries allows you to query multiple databases and S3 in a single query

A really useful feature in terms of a data platform is the ability to configure rules in Atlas that will trigger the automatic tiering or archiving of data from an Atlas cluster to an Atlas Data Lake. This moves the data targeted by the rules, typically older or infrequently accessed data from an Atlas cluster to a read-only Data Lake on S3 without requiring any user action, barring the initial rules setup. It also allows a unified view of both your Atlas and Data Lake data by querying the Data Lake.

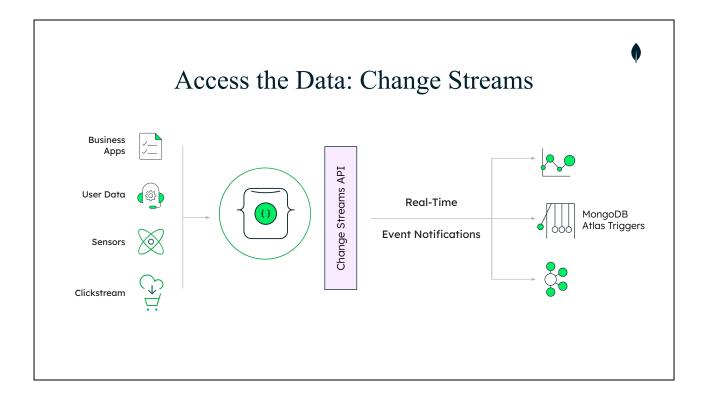
# Accessing the Data



Moving to the second aspect of a data platform which is accessing the data. The first place to start with is workload isolation.

Being able to run operational workloads along with analytical workloads on the same platform is a big advantage as it means you do not need to provision a separate database or set of resources to service your analytical queries. This is done by designating a specific set of nodes to be read-only, which will service the analytical queries.

MongoDB Atlas provides one set of nodes which are focused on servicing your main traffic for your operational applications. These nodes continue to replicate any changes or updates to a second set of nodes. This second set of nodes serves your analytics queries. Your application can target through which of the two sets of nodes should service the request depending on the purpose of the request.



MongoDB as a data platform has another useful feature but that is not as well known. Change streams provide a mechanism for applications to subscribe to changes whether on a single collection, a database, or an entire deployment. These are streaming as they happen to the application so it can react immediately to them.

Real-time data changes occuring in a range of applications, such as business applications, user data, sensor data, or clickstream data.

Change Streams and specifically the Change Stream API allows for applications to access and consume these real-time event notifications. Change streams aren't just limited to your application as they underpin a number of other tools in the wider MongoDB ecosystem. Change Streams use the aggregation framework so it allows for additional filtering or transformations to be applied to the changes as they are sent out.

MongoDB Charts can use Change Streams as a data source which can additionally be used with the Charts Embedded SDK to include the functionality of both directly into your application.

MongoDB Atlas triggers, specifically the database triggers, are powered by Change Streams. It allows the database triggers to listen for changes in watched collections and map them to database events. You can also run Realm functions on these events to further process them within in the MongoDB Data Platform.

MongoDB's Kafka Connector also uses Change Streams. It passes these notifications to Kafka so the details of the events can be stored in Kafka.

This functionality allows for reactivity to be built into your applications where any changes can be pushed from MongoDB to your application rather than your application having to regularly poll the database with additional queries to determine if there were any changes in terms of the data since the last time it checked.

Which of the following are true for MongoDB as a data platform? Select all that apply.

- A. Atlas Data Lake takes a variety of file formats
- B. Atlas Data Lake use MongoDB's Query Language (MQL)
- C. MongoDB does not support workload isolation
- D. Change Streams use the MongoDB Aggregation Framework

Which of the following are true for MongoDB as a data platform? Select all that apply.

- A. Atlas Data Lake takes a variety of file formats
- B. Atlas Data Lake use MongoDB's Query Language (MQL)
- C. MongoDB does not support workload isolation
- D. Change Streams use the MongoDB Aggregation Framework

CORRECT: Atlas Data Lake takes a variety of file formats - as well as JSON and BSON, it supports Avro, Parquet, ORC, CSV and TSV file formats.

CORRECT: Atlas Data Lake use MongoDB's Query Language (MQL) - You can query your data using the standard MQL, if you run a federated query you can query data both in your Atlas Cluster and in your Atlas Data Lake.

INCORRECT: MongoDB does not support workload isolation - Replica Set Tags are used in MongoDB within a replica set to support workload isolation.

CORRECT: Change Streams use the MongoDB Aggregation Framework - The Change Stream feature makes extensive use of the Aggregation Framework and it is what allows for changes to be made to the change stream event document prior to it being sent to a listening application.



Which of the following are true for MongoDB as a data platform? Select all that apply.

- A. Atlas Data Lake takes a variety of file formats
- B. Atlas Data Lake use MongoDB's Query Language (MQL)
- C. MongoDB does not support workload isolation
- D. Change Streams use the MongoDB Aggregation Framework

This is correct. Atlas Data Lake takes a variety of file formats - as well as JSON and BSON, it supports Avro, Parquet, ORC, CSV and TSV file formats.

CORRECT: Atlas Data Lake takes a variety of file formats - as well as JSON and BSON, it supports Avro, Parquet, ORC, CSV and TSV file formats.



Which of the following are true for MongoDB as a data platform? Select all that apply.

- A. Atlas Data Lake takes a variety of file formats
- B. Atlas Data Lake use MongoDB's Query Language (MQL)
- C. MongoDB does not support workload isolation
- D. Change Streams use the MongoDB Aggregation Framework

This is correct. You can query your data using the standard MQL, if you run a federated query you can query data both in your Atlas Cluster and in your Atlas Data Lake.

CORRECT: Atlas Data Lake use MongoDB's Query Language (MQL) - You can query your data using the standard MQL, if you run a federated query you can query data both in your Atlas Cluster and in your Atlas Data Lake.



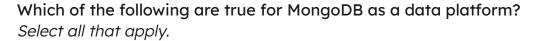
Which of the following are true for MongoDB as a data platform? Select all that apply.

- A. Atlas Data Lake takes a variety of file formats
- B. Atlas Data Lake use MongoDB's Query Language (MQL)
- C. MongoDB does not support workload isolation
- D. Change Streams use the MongoDB Aggregation Framework

This is incorrect. Replica Set Tags are used in MongoDB within a replica set to support workload isolation.

INCORRECT: MongoDB does not support workload isolation - Replica Set Tags are used in MongoDB within a replica set to support workload isolation.





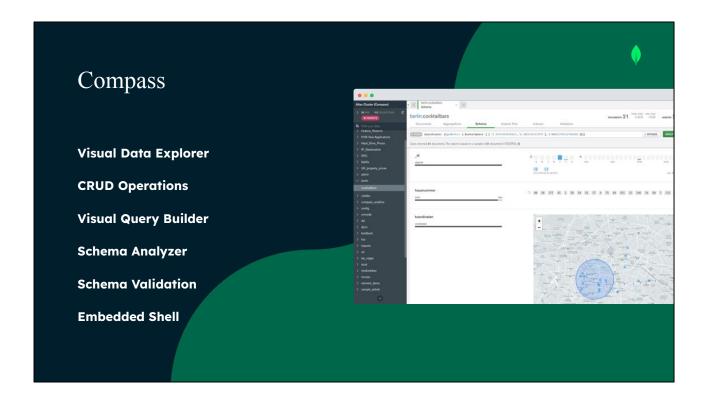
- A. Atlas Data Lake takes a variety of file formats
- B. Atlas Data Lake use MongoDB's Query Language (MQL)
- C. MongoDB does not support workload isolation
- D. Change Streams use the MongoDB Aggregation Framework

This is correct. The Change Stream feature makes extensive use of the Aggregation Framework.

CORRECT: Change Streams use the MongoDB Aggregation Framework - The Change Stream feature makes extensive use of the Aggregation Framework and it is what allows for changes to be made to the change stream event document prior to it being sent to a listening application.



# Processing the Data



Let's look at how the data can be processed.

The first tool within the data platform to introduce for process is MongoDB Compass, this is a visual data explorer. It runs as an application on your desktop or laptop and can connect to a local or to a remote MongoDB deployment. It lets you view the data with many different visualizations.

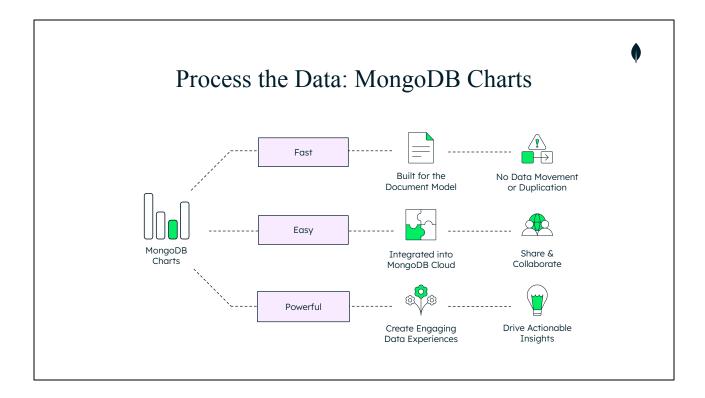
It allows for creating/reading/updating/deleting data or CRUD operations, on the documents in your collections.

It also includes visual query builders to help you construct complex queries for MongoDB Query Language and for MongoDB Aggregations.

MongoDB Compass also provides a schema analyzer which is the best way to identify schemas for a given dataset/datasets.

Schema validation supports JSON Schema format which you can edit/configure in Compass and look at live document previews of passing and of failing documents.

An embedded shell is also included in MongoDB Compass. This is a complete and functional Javascript environment that allows for interacting with MongoDB deployments. The MongoDB embedded shell allows for testing queries and for operations to be performed directly against your database within MongoDB Compass.



MongoDB Charts is the best way to visualize MongoDB data. With Charts...

- It's **fast** to visualize data
- It's easy to get started and share data
- And Charts is powerful to create engaging data experiences that drive actionable insights for internal and external users.

It is an analytical tool designed specifically for the document model, this provides it with additional flexibility when compared to using analytical tools designed for relational models with MongoDB data.

Charts versatility from Business Intelligence tool capabilities to embedded analytics with its quick and easy visualizations for the document model, help drastically **improve developer productivity.** 

- Data visualized in just minutes. Charts is built specifically for the document model, no ETL, no time loss to data manipulation or duplication required to visualize rich JSON data.
- Data shared quickly and easily. Integrated with Atlas as part of the MongoDB Cloud Data Platform, there's no setup for Charts. Go straight from Atlas into Charts to start sharing and collaborating on your MongoDB data.
- Powerful capabilities to create engaging data experiences for actionable insights for the right user in the right context. Whether via dashboard or a rich embedding SDK with authentication features, create engaging data-driven

•	experiences for everyone. From an internal dashboard to a customer-facing app, these experiences drive actionable insights across a wide spectrum of users.	

Which of the following are true for MongoDB as a data platform? Select all that apply.

- A. Compass provides a local tool to process MongoDB data
- B. Compass does not provide an embedded MongoDB Shell
- C. Charts does not allow for external embedding into other applications
- D. Business Intelligence Connector works with any ODBC compliant tool

Which of the following are true for MongoDB as a data platform? Select all that apply. More than 1 answer choice can be correct.

- A. Compass provides a local tool to process MongoDB data
- 8 B. Compass does not provide an embedded MongoDB Shell
- C. Charts does not allow for external embedding into other applications
- D. Business Intelligence Connector works with any ODBC compliant tool

CORRECT: Compass provides a local tool to process MongoDB data - Compass provides a GUI which you can use to visualise and connect to MongoDB deployment whether remote or local.

INCORRECT: Compass does not provide an embedded MongoDB Shell - Compass in recent versions does include an embedded MongoDB Shell/Javascript environment.

INCORRECT: Charts does not allow for external embedding into other applications - Charts does allow for a variety of external embedding whether using a iFrame or via its Javascript SDK for embedding data externally.

CORRECT: Business Intelligence Connector works with any ODBC compliant tool - The Business Intelligence Connector is designed to be usable with a range of tools as it interprets SQL, translates it to MQL and queries the MongoDB database, it then reformats the results back to SQL and sends these back to whatever ODBC compliant tool is connected.



Which of the following are true for MongoDB as a data platform? Select all that apply. More than 1 answer choice can be correct.

- A. Compass provides a local tool to process MongoDB data
- B. Compass does not provide an embedded MongoDB Shell
- C. Charts does not allow for external embedding into other applications
- D. Business Intelligence Connector works with any ODBC compliant tool

This is correct. Compass provides a GUI which you can use to visualize and connect to MongoDB deployment whether remote or local.

CORRECT: Compass provides a local tool to process MongoDB data - Compass provides a GUI which you can use to visualise and connect to MongoDB deployment whether remote or local.

Which of the following are true for MongoDB as a data platform? Select all that apply. More than 1 answer choice can be correct.

- A. Compass provides a local tool to process MongoDB data
- B. Compass does not provide an embedded MongoDB Shell
- C. Charts does not allow for external embedding into other applications
- D. Business Intelligence Connector works with any ODBC compliant tool

This is incorrect.

Compass in recent versions does include an embedded MongoDB

Shell/Javascript environment.

INCORRECT: Compass does not provide an embedded MongoDB Shell - Compass in recent versions does include an embedded MongoDB Shell/Javascript environment.



Which of the following are true for MongoDB as a data platform? Select all that apply. More than 1 answer choice can be correct.

- A. Compass provides a local tool to process MongoDB data
- B. Compass does not provide an embedded MongoDB Shell
- C. Charts does not allow for external embedding into other applications
- D. Business Intelligence Connector works with any ODBC compliant tool

This is incorrect. Charts does allow for a variety of external embedding whether using a iFrame or via its Javascript SDK for embedding data externally.

INCORRECT: Charts does not allow for external embedding into other applications - Charts does allow for a variety of external embedding whether using a iFrame or via its Javascript SDK for embedding data externally.



Which of the following are true for MongoDB as a data platform? Select all that apply. More than 1 answer choice can be correct.

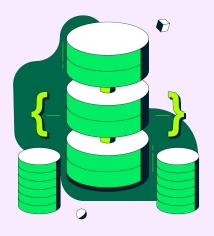
- A. Compass provides a local tool to process MongoDB data
- B. Compass does not provide an embedded MongoDB Shell
- C. Charts does not allow for external embedding into other applications
- D. Business Intelligence Connector works with any ODBC compliant tool

This is correct. It is usable with a range of tools as it interprets SQL, translates it to MQL and queries the MongoDB database, it then reformats the results back to SQL and then back to the ODBC compliant tool.

CORRECT: BI Connector works with any ODBC compliant tool - It is usable with a range of tools as it interprets SQL, translates it to MQL and queries the MongoDB database, it then reformats the results back to SQL and then back to the ODBC compliant tool.







Storing the data:
Self-Managed, Atlas, and
Atlas Data Lake

Accessing the data: Workload isolation and Change Streams

Processing the data:
Compass, Charts, and the
Business Intelligence
Connector

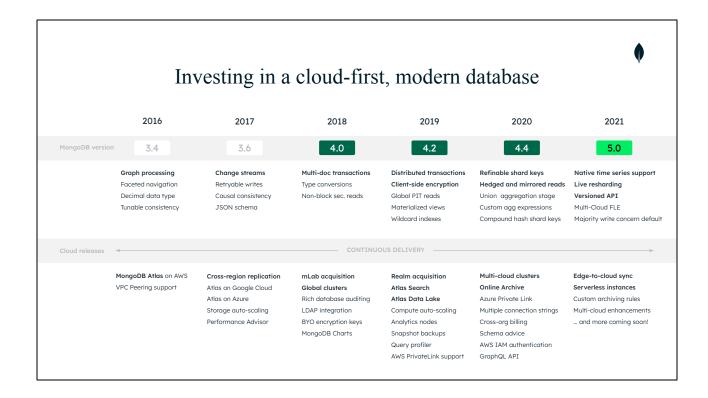
We started this lesson highlighting that MongoDB is not just a database product, it is a complete ecosystem.

We pointed out that it is better viewed as a data platform with a number of tools that help with various aspects.

In storing the data, we looked at self-managed MongoDB, MongoDB Atlas which is the SaaS or DBaaS offering from MongoDB and includes features such as Atlas Search. We also introduced Atlas Data Lake which provides a data lake facility allowing you to separate archival data from your operational workload and the data it requires.



Let's focus in the next section on how MongoDB Cloud is the foundation and underpinning of the MongoDB Data Platform. MongoDB Cloud encompasses the pieces of the data platform that are provided in a SaaS manner and where all the operational plus maintenance aspects are managed completely by MongoDB for the user.



We've invested as much if not more in building out our cloud products and capabilities alongside continuous innovation on the core database software.

A major benefit of taking a cloud-first approach is the ability to change to a continuous delivery model with a cycle of three weeks rather than the prior annual release cycle for the software. This means users see new features far sooner and it allows for these to be regularly pushed out, further improving the developer experience.





Global and Multi-cloud Reach



Performance Advisory



Online Archive



**Atlas Search** 

MongoDB Cloud platform offers true multi-cloud capabilities as well as global capacities. It bakes in automatic instrumentation with performance advisory to dynamically monitor your workloads and make actionable index suggestions. In terms of easily archiving data, the online archive feature allows for colder storage to be used with defined rules to migrate data from your production clusters to cold storage/analytical storage (typically S3). Atlas Search allows for any application to easily add search functionality backed by Lucene via Atlas without having to stand up and maintain a separate search cluster, this means your data can be stored, used and searched in Atlas easily. Lucene allows for highly fine grained searches to be applied on your MongoDB database.

## A Cloud-First, Modern Database



**Atlas Charts** 







Global Clusters

Continuous Cloud Backup **Atlas Serverless** 

Atlas Charts allows for rapid data visualization within the environment without having to export to another separate data visualization tool. Global clusters allows for all users and applications regardless of their geographical location to have the optimal experience and latency, not to mention easily managing any kind of data compliance such as Europe's GDPR. In terms of backup, Atlas offers a continuous backup feature which allows you to define your own backup and retention policies, this can satisfy a recovery point objective as low as 60 seconds. Atlas Serverless offers a option for developers using the Serverless paradigm to use MongoDB's Atlas Serverless instances to easily scale up and down to their immediate needs of the application/traffic.







Realm Sync





Data Federation

Kubernetes Operator

**BI** Connector

Data Federation provides a query engine that allows for the federation of queries across a variety of different sources of data as well as the ability to move data between these sources. Realm Sync is a edge-to-cloud data synchronization service which simplifies keeping devices which aren't always online synced with your database by handling the conflict resolution and networking code. MongoDB's Atlas Operator allows you to simplify deployment, management and scaling of your Atlas clusters in Kubernetes. The BI Connector allows for a wide range of analytical tooling designed for relational/SQL databases to easily use MongoDB databases so you can use advanced visualization from the likes of Tableau to represent and interpret your data.

Which of the following are true for MongoDB Cloud within the data platform? Select all that apply.

- A. MongoDB Atlas Search provides granular text searches
- B. Federated Data cannot be queried, data can only be archived to it
- C. Realm Sync is designed for edge devices
- D. MongoDB Charts only uses data from MongoDB Atlas collections

Which of the following are true for MongoDB Cloud within the data platform? Select all that apply.

- A. MongoDB Atlas Search provides granular text searches
- B. Federated Data cannot be queried, data can only be archived to it
- C. Realm Sync is designed for edge devices
- D. MongoDB Charts only uses data from MongoDB Atlas collections

CORRECT: MongoDB Atlas Search provides granular text searches - MongoDB Atlas Search utilizes Apache Lucene to provide granular text searches. This service also manages automatically keeping the indexes updated as data changes. INCORRECT: Federated Data cannot be queried, data can only be archived to it - Federated Data can be queried, it's purpose is to be both an archive as well providing the ability to query the data present there. It does not support indexes. CORRECT: Realm Sync is designed for edge devices - Realm is designed for edge devices and offers features like data sync so data serialization, network recovery, and conflict resolution are all handled by Realm itself rather than by your application. CORRECT: MongoDB Charts only uses data from MongoDB Atlas collections - Charts are designed to work directly from MongoDB Atlas databases and collections, it is not a standalone product but sits within MongoDB Cloud as a feature (visualization).

Which of the following are true for MongoDB Cloud within the data platform? Select all that apply.

- A. MongoDB Atlas Search provides granular text This is correct. MongoDB searches
- B. Federated Data cannot be queried, data can only be archived to it
- C. Realm Sync is designed for edge devices
- D. MongoDB Charts only uses data from MongoDB Atlas collections

This is correct. MongoDE Atlas Search utilizes Apache Lucene to provide granular text searches. This service also manages automatically keeping the indexes updated as data changes.

CORRECT: MongoDB Atlas Search provides granular text searches - MongoDB Atlas Search utilizes Apache Lucene to provide granular text searches. This service also manages automatically keeping the indexes updated as data changes.



Which of the following are true for MongoDB Cloud within the data platform? Select all that apply.

- A. MongoDB Atlas Search provides granular text
   Searches

  Fodorated Data
- B. Federated Data cannot be queried, data can only be archived to it
- ✓ C. Realm Sync is designed for edge devices
- D. MongoDB Charts only uses data from MongoDB Atlas collections

This is incorrect.
Federated Data can be queried, it's purpose is to be both an archive as well providing the ability to query the data present there. It does not support indexes.

INCORRECT: Federated Data cannot be queried, data can only be archived to it - Federated Data can be queried, it's purpose is to be both an archive as well providing the ability to query the data present there. It does not support indexes.



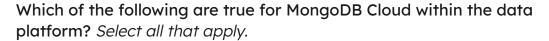
Which of the following are true for MongoDB Cloud within the data platform? Select all that apply.

- A. MongoDB Atlas Search provides granular text This is correct. Realm is searches
- B. Federated Data cannot be queried, data can only be archived to it
- C. Realm Sync is designed for edge devices
- D. MongoDB Charts only uses data from MongoDB Atlas collections

This is correct. Realm is designed for edge devices and offers features like data sync so data serialization, network recovery, and conflict resolution are all handled by Realm itself rather than by your application.

CORRECT: Realm Sync is designed for edge devices - Realm is designed for edge devices and offers features like data sync so data serialization, network recovery, and conflict resolution are all handled by Realm itself rather than by your application.



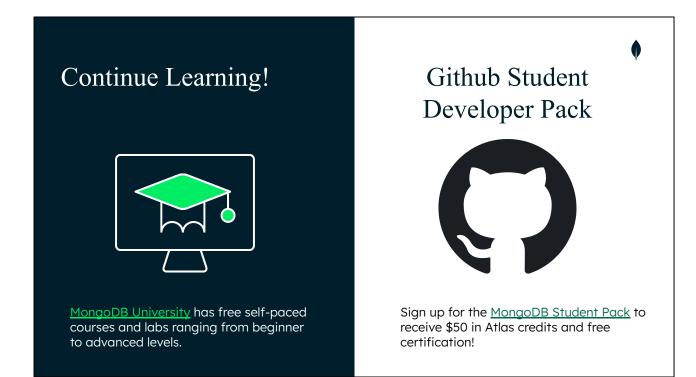


- A. MongoDB Atlas Search provides granular text This is correct. Charts are searches
- B. Federated Data cannot be queried, data can only be archived to it
- C. Realm Sync is designed for edge devices
- D. MongoDB Charts only uses data from MongoDB Atlas collections

This is correct. Charts are designed to work directly from MongoDB Atlas databases and collections, it is not a standalone product but sits within MongoDB Cloud as a feature (visualization).

CORRECT: MongoDB Charts only uses data from MongoDB Atlas collections - Charts are designed to work directly from MongoDB Atlas databases and collections, it is not a standalone product but sits within MongoDB Cloud as a feature (visualization).





This concludes the material for this lesson. However, there are many more ways to learn about MongoDB and non-relational databases, and they are all free! Check out MongoDB's University page to find free courses that go into more depth about everything MongoDB and non-relational. For students and educators alike, MongoDB for Academia is here to offer support in many forms. Check out our educator resources and join the Educator Community. Students can receive \$50 in Atlas credits and free certification through the Github Student Developer Pack.