MongoDB Atlas

Google slide deck available [here](#)

Overview

Learning Objectives

At the end of this lesson, learners will be able to:
- Explain the architecture of MongoDB's multi-cloud product Atlas.
- Define Database as a Service.
- Explain the difference between a multi-cloud platform and cross-cloud platform.
- Identify the functions of Atlas' clusters and connectors to ease the movement of data.
- Identify the eight key features of Atlas.

Suggested Uses

- Lecture for one hour class or a part of a longer lecture period
- Handouts / asynchronous learning
- Supplemental reading material - read on your own / not part of formal teaching
- Complement to University courses M001: MongoDB Basics.

This lesson is a part of the courses MongoDB: A Developer Data Platform and Introduction to Modern Databases with MongoDB.

At a Glance

Length:
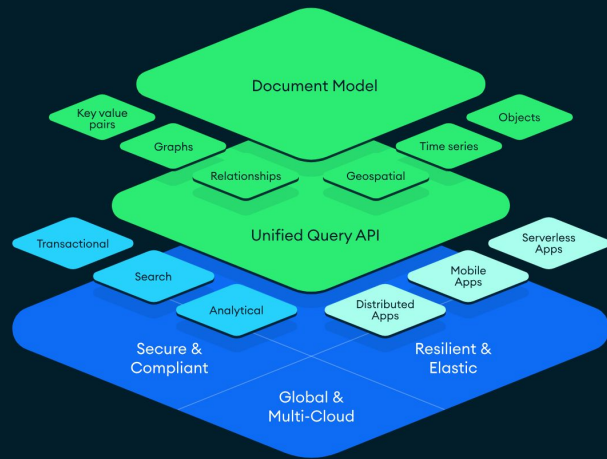45 minutes

Level:
Foundational

Prerequisites:
None

Share your feedback: We hope these curriculum materials will be a valuable resource for you and your learners. Let us know how the materials work for you, what we can improve on, and how MongoDB for Academia can support you via our brief feedback form.

MongoDB for Academia: MongoDB for Academia offers resources for educators and students to support teaching and learning MongoDB. 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.

Last Update: July 2022

Now, let's focus on the foundation of resilience, scale, and security at a very deep level of the MongoDB architecture.

MongoDB Atlas: Global Cloud Database

Intelligent performance optimization

Simplified data architecture

Best-in-class security and operations
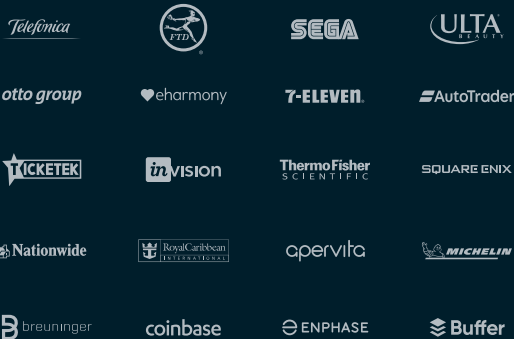
Global and multi-cloud reach

80+ Regions

Building on MongoDB's intuitive data model and query API, MongoDB Atlas gives engineering organizations the versatility they need to build sophisticated applications that can adapt to changing customer demands and market trends. Not only is it the only multi-cloud document database available, it also delivers the most advanced security and data distribution capabilities of any fully managed service. Get started in minutes and leverage intelligent automation to maintain performance at scale as your applications evolve over time.

What makes MongoDB Atlas the best database service for your teams?
- Focus on consistently shipping exceptional features and products with MongoDB Atlas' intelligent performance optimization tools
- Reduce complexity across systems with a simplified data architecture and API
- Scale your teams confidently with MongoDB Atlas' best-in-class infrastructure operations
- Run anywhere in the world with MongoDB Atlas' multi-cloud reach

MongoDB Atlas is a Database as a Service (DBaaS), this is similar to Software as a Service (SaaS).

SaaS is where software is hosted centrally and you as a consumer/user of the software pays a subscription to use it. You don't have to run it on your machines as the SaaS provider deals with the hosting, maintenance and general running of the software. In the database space,

SaaS software that is focused on the provision of database and related hosting is often called DBaaS or Database as a Service.

Forrester names MongoDB a leader in the Database as a Service market due to MongoDB Atlas. This puts in the same category as Amazon, Microsoft, Oracle, and Google.
The difference is that we're in the leader category because of one product, MongoDB Atlas. The other companies categorized as leaders got there because of multiple databases. This shows that Atlas is a comprehensive database service that stands above and beyond the other solutions you'll find in the market.

Multi-cloud clusters on MongoDB Atlas

Industry first and unique

Gives you unparalleled flexibility when it comes to where your data is stored and what cloud services you can use with MongoDB

- Take advantage of best-of-breed technology across multiple clouds

- Seamlessly migrate your cluster from one cloud to another

- Improve high availability with cross-cloud redundancy

- Reach more users by distributing your database across more regions

Google Cloud   aws   Azure

Multiple clouds simultaneously

MongoDB Atlas was the first database to offer DBaaS across multiple cloud providers where deployments could be created using two or more different public cloud providers. This helps in ensuring that deployments can offer cross-cloud (provider) redundancy.

Let's look next at what the difference is between cross-cloud and multi-cloud.

Multi-Cloud

Avoid vendor lock-in

Capitalize on the strengths of each

Ability to move (relocate)

Easily migrate an application, or a tenant (data)

Use data stored in different clouds to power a single application, or vice-versa.

Cross-Cloud

Provide business continuity under natural disasters and/or man-made (attacks)

Meet regulations within a geography irrespective of cloud provider.

A cross-cloud capability is necessary to enable secure data sharing across cloud providers and regions.

Capitalize on the strengths of each without manually managing data movement.

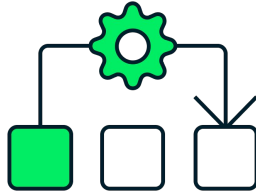Meet customers where they are irrespective of cloud provider (e.g. gaming)

Key technologies that enable each

Stay close to OSS vs. Cloud adaptations
Kafka vs. Cloud PubSub
ElasticSearch vs.
K8S, OSB vs. GKE, EKS or AKS

Database triggers
Automate function execution in response to database changes or user events

Connectors
Seamlessly integrate with Kafka, Spark, and business intelligence tools

GraphQL API
Automatically generate a JSON schema and enable GraphQL for apps with a single click

Ease of data access and movement

There are a number of features in Atlas that support data access and movement and later in this lessons we'll discuss Atlas Search, Atlas Data Federation, and the Data Explorer. We'll just briefly cover database triggers, connectors, and Atlas's GraphQL API in this slide before moving to a brief introduction on how Atlas works as a DBaaS.

With Triggers in MongoDB Atlas it's seamless to create scalable responses to database events or scheduled workloads.

The MongoDB Connector for Apache Spark, enables organizations to fully realize the potential of real-time analytics. Spark jobs can be executed directly against operational data managed by MongoDB, without the time and expense of Extract Transform Load (ETL) processes. MongoDB can efficiently index and serve analytics results back into live, operational processes, making them smarter, more contextual and responsive to events as they happen.

The MongoDB Connector for Apache Kafka allows you to easily build robust and reactive data pipelines that take advantage of stream processing between datastores, applications and services in real-time. Eliminate the need for writing boilerplate integration code and ensure end-to-end data exchange between any source and sink.

MongoDB's GraphQL service makes it easy to get up and running with GraphQL, the groundbreaking API query language that gives developers a single endpoint to access exactly the data they need - nothing more, nothing less.
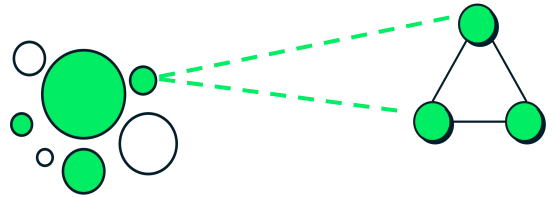
How Atlas works: Clusters

Clusters

A group of servers that
stores your data

In MongoDB Atlas, clusters are the unit of deployment.
Each cluster represents either a replica set or a sharded cluster.
Each cluster is the group of servers that holds your data.

Clusters: each Atlas cluster can be a replica set or it can be a sharded cluster.

In MongoDB Atlas, clusters are the unit of deployment.
Each cluster represents either a replica set or a sharded cluster.
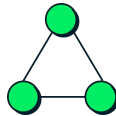Each cluster is the group of servers that holds your data.

What is an instance?

Clusters

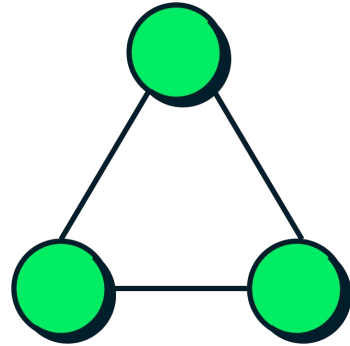A group of servers that stores your data

Replica set

A few connected MongoDB instances that store the same data

A replica set in Atlas is made up of three or more mongodb instances that store the same data.

In Atlas, an instance is defined as any single machine whether in the cloud or running locally running the specific software (in this case the MongoDB database server).

Instance: a single machine locally or in the cloud, running a certain software

A replica set in Atlas is made up of three or more mongodb instances that store the same data.

In Atlas, an instance is defined as any single machine whether in the cloud or running locally the specific software (in this case the MongoDB database server).
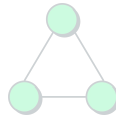
Atlas as a DBaaS

Clusters	Replica set	Atlas
A group of servers that stores your data	A few connected MongoDB instances that store the same data	Database as a service

Atlas is the software which runs all the scaffolding to manage all of the clusters of all the customers as well as providing the additional services/features that we'll cover in the coming slides.

It manages all the hosting aspects for you as an end user and at its simplest provides you with the connection string required by your driver/application to connect to the database and use it.

MongoDB Atlas uses all three of the major public cloud providers (Amazon, Google, and Microsoft) and a variety of geographical regions within each of these three cloud providers.

Quiz

Quiz

Which of the following are true for MongoDB Atlas? *Select all that apply.*

- A. You host / run it on your hardware
- B. It provides Database as a Service using cloud providers
- C. A replica set is the basic building block for clusters
- D. It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data

Quiz

Which of the following are true for MongoDB Atlas? *Select all that apply.*

❌ A. You host / run it on your hardware

✅ B. It provides Database as a Service using cloud providers

✅ C. A replica set is the basic building block for clusters

✅ D. It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data

INCORRECT: You host/run it on your hardware - MongoDB Atlas is hosted by MongoDB and run on public clouds not on your hardware
CORRECT: It provides database as a service using cloud providers - MongoDB Atlas uses AWS, GCP, and Azure to provide the hosting for its Database as a Service
CORRECT: A replica set is the basic building block for clusters - The cluster is the atomic unit of organisation in MongoDB Atlas, within each cluster the basic building block is a replica set.
CORRECT: It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data - Atlas provide a range of additional services and the next lesson will explore these in more depth.

Quiz

Which of the following are true for MongoDB Atlas? *Select all that apply.*

❌ A. You host / run it on your hardware

✅ B. It provides Database as a Service using cloud providers

✅ C. A replica set is the basic building block for clusters

✅ D. It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data

This is incorrect. MongoDB Atlas is hosted by MongoDB and runs on public clouds not on your hardware.

INCORRECT: You host/run it on your hardware - MongoDB Atlas is hosted by MongoDB and runs on public clouds not on your hardware

Quiz

Which of the following are true for MongoDB Atlas? *Select all that apply.*

❌ A. You host / run it on your hardware

✅ B. It provides Database as a Service using cloud providers

✅ C. A replica set is the basic building block for clusters

✅ D. It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data

This is correct. MongoDB Atlas uses AWS, GCP, and Azure to provide the hosting for its Database as a Service.

CORRECT: It provides database as a service using cloud providers - MongoDB Atlas uses AWS, GCP, and Azure to provide the hosting for its Database as a Service

Quiz

Which of the following are true for MongoDB Atlas? *Select all that apply.*

❌ A. You host / run it on your hardware

✅ B. It provides Database as a Service using cloud providers

✅ C. A replica set is the basic building block for clusters

✅ D. It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data

This is correct. The cluster is the atomic unit of organization in MongoDB Atlas, within each cluster the basic building block is a replica set.

CORRECT: A replica set is the basic building block for clusters - The cluster is the atomic unit of organization in MongoDB Atlas, within each cluster the basic building block is a replica set.

Quiz

Which of the following are true for MongoDB Atlas? *Select all that apply.*

❌ A. You host / run it on your hardware

✅ B. It provides Database as a Service using cloud providers

✅ C. A replica set is the basic building block for clusters

✅ D. It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data

This is correct. Atlas provides a range of additional services and the next part of the lesson will explore these in more depth.

CORRECT: It has additional services for data explorations, scheduled jobs/work, data federation, and full text searching of data - Atlas provides a range of additional services and the next part of the lesson will explore these in more depth.
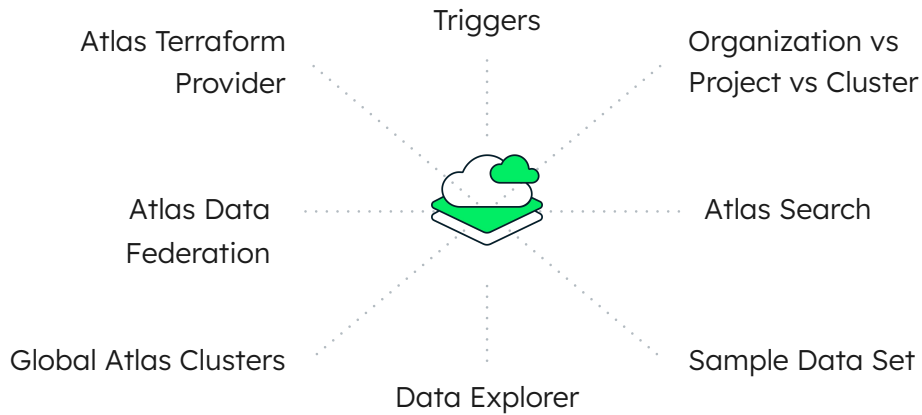
MongoDB Atlas

Let's focus in the next section on a short tour of some of the features in MongoDB Atlas, it isn't designed to be a complete feature list, rather this is just a sample to help highlight the flexibility and range of functionality available in Atlas.

Atlas Short Feature Tour

Triggers

Atlas Terraform Provider

Organization vs Project vs Cluster

Atlas Data Federation

Atlas Search

Global Atlas Clusters

Sample Data Set

Data Explorer

In this part of the lesson, we'll look at each of these features in a little more detail to help you understand exactly what MongoDB Atlas provides beyond hosting and maintenance of your database in the cloud.

Atlas Sample
Data Set

Atlas provides
nine sample
datasets to help
you investigate
and explore ideas.

Firstly, let's look at the Atlas Sample Data Set.

The full list of nine datasets are Sample AirBnB Listings Dataset, Sample Analytics
Dataset, Samples Guides, Sample Geospatial Dataset, Sample Mflix Dataset, Sample
Restaurants Dataset, Sample Supply Store Dataset, Sample Training Dataset, and
the Sample Weather Dataset.

These cover everything from AirBnB listings to Movie and Restaurant reviews. They
provide a broad spectrum of data which you can use to prototype any ideas without
first needing to have to create your own data.

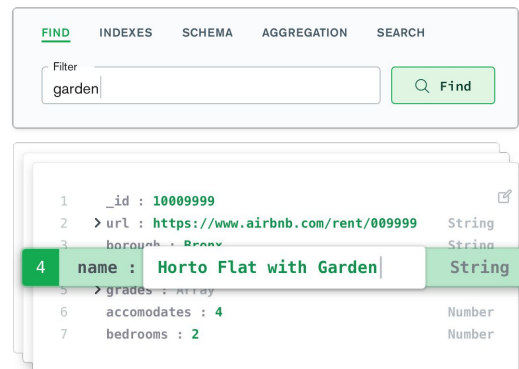Each of these has documentation around their specific schema and indexes which
can be found at
https://docs.atlas.mongodb.com/sample-data/available-sample-datasets/

This built-in tool to allow you interact with data

CRUD operations on documents, collections, and databases

Manage indexes

Create & run aggregations

FIND   INDEXES   SCHEMA   AGGREGATION   SEARCH

Filter
garden   🔍 Find

```
1     _id : 10009999
2   > url : https://www.airbnb.com/rent/009999    String
3     borough : Bronx                              String
4     name :   Horto Flat with Garden            String
5   > grades : Array
6     accomodates : 4                             Number
7     bedrooms : 2                                Number
```

Atlas Data Explorer

The second of Atlas's features that we will investigate is the Atlas Data Explorer.

Useful UI feature that supports the majority of common operations you need to interact with your data via your web browser.
Atlas's Data Explorer allows you to easily perform all of these operations within the Atlas UI without any additional tools other than your browser.

Specifically the Atlas Data Explorer allows you to do the following:
- Create, Read, Update, and Delete (CRUD) operations on documents, collections, and databases
- Manage indexes
- Create and run aggregations

User authorization can be set for all of these tasks to ensure user permissions can be applied if required.

Unifies data across clusters into Virtual Databases and Collections

Allows you to query it as normal using MQL

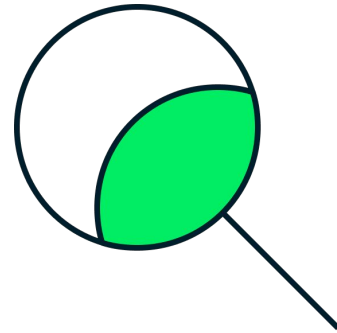The data stays where it is, no copies or moves

Atlas Data Federation

Thirdly, we'll look at the Atlas Data Federation, which was designed to make it easier for people to query their data no matter where it is in their deployment.

Data Federation allows you to create virtual databases and collections from your existing data which can be a combination of any set of existing databases and existing collections. The important thing is that these virtual instances do not change your data or cause it to be copied or moved in any fashion.

A query in Data Federation will be directed to where the real data lives so you can easily create virtual databases to service applications but without having to copy or move data into a new database or write queries that target multiple databases/collections. This greatly helps to simplify scenarios which require this type of querying.

Fuzzy search

Autocomplete

Filters and facets

Multi-data type support

Multi-language support

Highlighting

Custom scoring

Rich query language

Atlas Search

The fourth feature we'll explore is Atlas Search.

MongoDB already has text indexes but Atlas Search leverages Apache Lucene under the hood to bring even more powerful search features to the database in Atlas. An example of where this fits is when you want functionality to allow a type-ahead with autocomplete in your application.

Specifically Atlas Search provides the following:
- Advanced search functionality without any additional management required, the fine grained indexing allows for certain fields to be indexed be only a specific analyzer or defining that multiple analyzers can index the field. Additionally, you can exclude certain fields or alternatively dynamically index all the fields in a collection.
- Analyzers control the way your full-text index is constructed and uses the $search aggregation pipeline stage. An analyzer provides control over the vocabulary and grammar. This allows for the optimal use of a particular language's vocabulary and grammar or alternative you can use it to create an index which is explicitly language neutral.
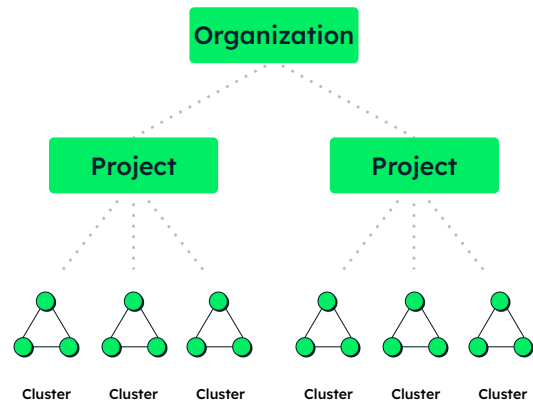
Atlas Organizations and Projects

Organizations relate to billing

Projects are best associated to teams

Each project has one or more cluster.

Applications connect to clusters

Organization

Project

Project

Cluster

Cluster

Cluster

Cluster

Cluster

Cluster

The fifth feature we will investigate relates to two concepts that are often confused in Atlas's features, Atlas organizations and Atlas projects.

We'll walk through this slide to highlight the concepts to clarify any potential confusion about the differences between organizations and projects in Atlas.
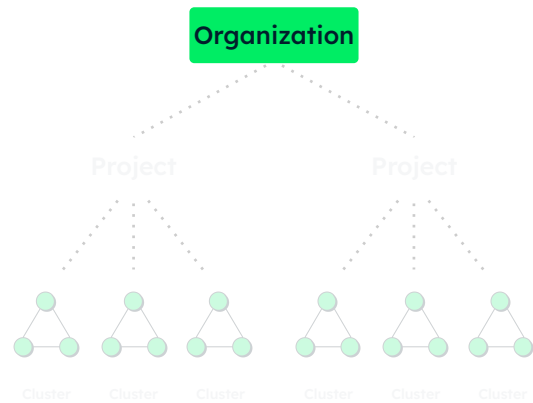
Atlas Organizations and Projects

Organizations relate to billing

Projects are best associated to teams

Each project has one or more clusters.

Applications connect to clusters

The easiest way to view these is that billing relates to an organization (for your company or organization typically) and projects related to teams within the organization. Within a project there can be clusters for various different environments such as development, staging, testing, production, etc.
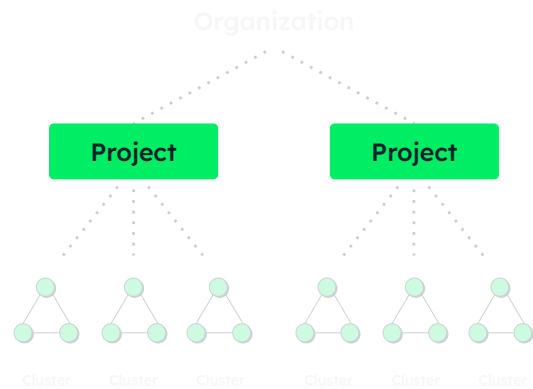
Projects are best associated to teams as this helps ensure there is a clear mapping between environments (at the cluster level) and their owners (best done at the project level) and the total billing (done per project at the organization level).

Additionally, a number of Atlas features such as Atlas Data Lake and Triggers are done at a project level.
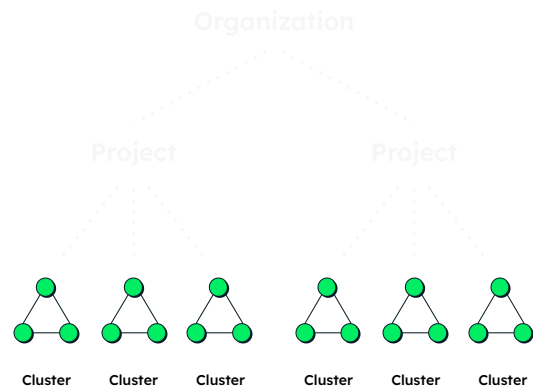
Atlas project's typically have one or more clusters. The multiple clusters refer to the different environments for a given application or at a coarser level than this (e.g. a cluster per application). Different environments are typically used for developing, staging, and production. There may be multiple layers in each environment, each layer typically corresponding or associated to a specific Atlas cluster.

Applications (typically via MongoDB Drivers) connect to specific Atlas clusters. As mentioned previously, the cluster is the atomic unit of deployment and organization within MongoDB Atlas.

Responds to events or on a schedule

Trigger types

Database

Scheduled

The sixth feature we'll investigate is Atlas Triggers. This is server-side code that can respond to database events or according to a schedule.

Atlas has two kinds of triggers, database and scheduled.

A database trigger will execute server-side whenever a document is added, updated or removed in a cluster that is linked.
Database triggers use MongoDB change streams to listen for changes in watched collections and map these to database events. We'll cover MongoDB change streams in more depth in a later lesson.

Scheduled triggers are very similar to traditional Unix/Linux CRON jobs, these are defined using the CRON temporal expressions to indicate when they should run. A CRON expression consists of five space-delimited fields: minute, hour, dayOfMonth, month, weekDay. These can be set to a specific value or to an expression that evaluates to a set of values. Overall, these are quite similar to CRON jobs in Unix/Linux and this was a deliberate design choice to simplify the use of this feature but using the well-established CRON format and syntax.

The schedule trigger will then execute the server-side logic on that CRON schedule. This type of trigger functionality is suited to periodic work, such as nightly reporting, updating documents every minute/hour, or indeed generating weekly newsletters or offer emails.
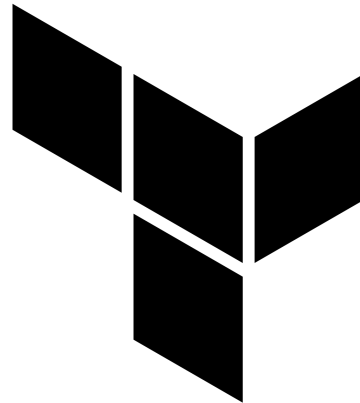
Uses declarative config files

Reproducible environments

Provisioning to majority of public cloud providers

Infrastructure as Code

Atlas and Hashicorp Terraform

The seventh feature we will explore is the MongoDB Atlas provider that was jointly developed by Hashicorp and MongoDB. It is used to interact and deploy MongoDB Atlas.

Terraform is a provisioning tool rather than a configuration management tool. This distinction is one of the key aspects of Terraform when compared to other infrastructure as code tools such as Chef, Puppet, or Ansible, all of these other tools focus on configuration management as their main functionality.
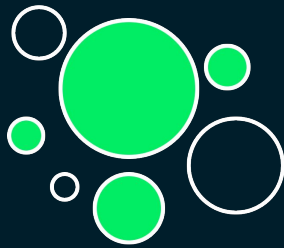
The full description and details on the MongoDB Atlas Terraform Provider can be found at the website on the slide https://www.mongodb.com/atlas/hashicorp-terraform

A provider in Terraform is used to create, manage, and update infrastructure resources. It is able to understand both Terraform's API and the provider's APIs (in this case MongoDB Atlas's API) and expose the provider's resources.

Specifically you should note the following about Terraform:
- It uses a declarative config file, where the described goal is declared rather than the steps or sequences of how to get to that goal (imperative configuration).
- It allows for the reproduction of the configuration so it is easy to generate the same environment again without any additional steps or configuration.
- Provisions to the vast majority of cloud providers, whether major or minor player in the field, and regardless of whether infrastructure, platform or

- software (as a service) providers (IaaS, PaaS, or SaaS).
- Describes your infrastructure using the declarative configuration, this allows your deployments to be versioned and treated in a similar fashion to how you manage software source code. It also allows these configurations to be easily shared and re-used.

Curated implementation of zone sharding to support:

Location-aware read and write operations

Application instances + clients can be globally distributed

Atlas Global Clusters utilize a lot of MongoDB know-how and technical knowledge to provide a curated implementation where shards are spread globally.

Atlas Global Clusters were designed to support the following:

- Location-aware data storage in specific geographic regions. This provides low-latency read and write operations for globally distributed clients.
- Locating your application instances and clients globally

Another advantage of a global cluster is a protection from a complete outage during partial or full regional outages.

Quiz

Quiz

Which additional features or services does MongoDB Atlas provide outside of managing your database? *Select all that apply.*

● A. Sample datasets

● B. Fine-grained text indexing

● C. Integration with the configuration management tools, specifically with Puppet, or with Chef, or with Ansible

Quiz

Which additional features or services does MongoDB Atlas provide outside of managing your database? *Select all that apply.*

- ✅ A. Sample datasets
- ✅ B. Fine-grained text indexing
- ❌ C. Integration with the configuration management tools, specifically with Puppet, or with Chef, or with Ansible

CORRECT: Sample datasets - Atlas provides 9 datasets that you can use immediately with Atlas to explore the platform and potentially prototype your applications with.

CORRECT: Fine grained text indexing - Atlas Search offers more comprehensive search with finer granularity than the default text indexes available in the MongoDB server. This is because Atlas Search uses Lucene as it's backend and this offers a much greater degree of functionality and control when creating indexes for text.

INCORRECT: Integration with Puppet, Chef, and Ansible - Atlas only offers integration with Hashicorp's Terraform and whilst there are community support modules for Puppet, Chef, and Ansible these are not developed in conjunction with MongoDB.

Quiz

Which additional features or services does MongoDB Atlas provide outside of managing your database? *Select all that apply.*

✅ A. Sample datasets

✅ B. Fine-grained text indexing

❌ C. Integration with the configuration management tools, specifically with Puppet, or with Chef, or with Ansible

This is correct. Atlas provides nine datasets that you can use immediately with Atlas to explore the platform and potentially prototype your applications with.

CORRECT: Sample datasets - Atlas provides nine datasets that you can use immediately with Atlas to explore the platform and potentially prototype your applications with.

Quiz

Which additional features or services does MongoDB Atlas provide outside of managing your database? *Select all that apply.*

✅ A. Sample datasets

✅ B. Fine-grained text indexing

❌ C. Integration with the configuration management tools, specifically with Puppet, or with Chef, or with Ansible

This is correct. Atlas Search offers more comprehensive search with finer granularity than the default text indexes available in the MongoDB server. This is because Atlas Search uses Lucene as the backend.

CORRECT: Fine grained text indexing - Atlas Search offers more comprehensive search with finer granularity than the default text indexes available in the MongoDB server. This is because Atlas Search uses Lucene as the backend.

Quiz

Which additional features or services does MongoDB Atlas provide outside of managing your database? *Select all that apply.*

✅ A. Sample datasets

✅ B. Fine-grained text indexing

❌ C. Integration with the configuration management tools, specifically with Puppet, or with Chef, or with Ansible

This is incorrect. MongoDB Atlas offers a Hashicorp Terraform integration but it does not integrate yet with these other tools.

INCORRECT:Integration with the configuration management tools, specifically with Puppet, or with Chef, or with Ansible - MongoDB Atlas offers a Hashicorp Terraform integration but it does not integrate yet with these other tools.

Continue Learning!

GitHub Student Developer Pack

MongoDB University has free self-paced courses and labs ranging from beginner to advanced levels.

Sign up for the MongoDB Student Pack to receive $50 in Atlas credits and free certification!

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.