Document Database

CS 4440

Intro

What is Document Database?

- A database that stores information in documents + NoSQL.

- Store data in flexible, schema-less format such as JSON, BSON, or XML (Rathore & Bagui, 2024).

```json
{
    "_id": "5cf0029caff5056591b0ce7d",
    "firstname": "Jane",
    "lastname": "Wu",
    "address": {
        "street": "1 Circle Rd",
        "city": "Los Angeles",
        "state": "CA",
        "zip": "90404"
    },
    "hobbies": ["surfing", "coding"]
}
```

Source: JSON Format of MongoDB

What is Document Database?

- Ideal for apps need rapid iteration, scalability, handling of hierarchical or nested data structures.

- E.g.: management systems, real-time analytics, e-commerce platforms, mobile apps
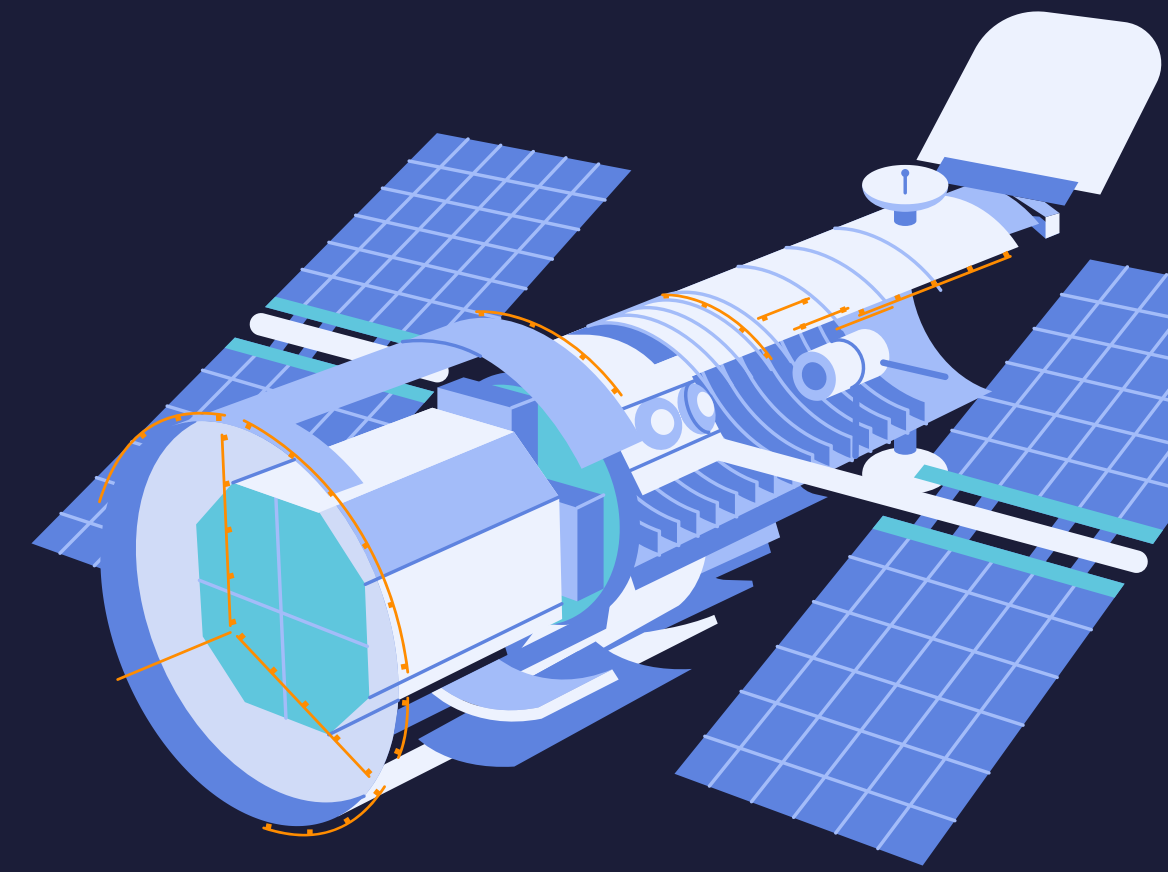
LMS

Mobile app

E-commerce

Features/ Functionalities

- Document Model
  - Attributes and values
- Flexible Schema
  - Optional attributes
  - Schema Validation
- CRUD operations
  - create, retrieval, update, delete
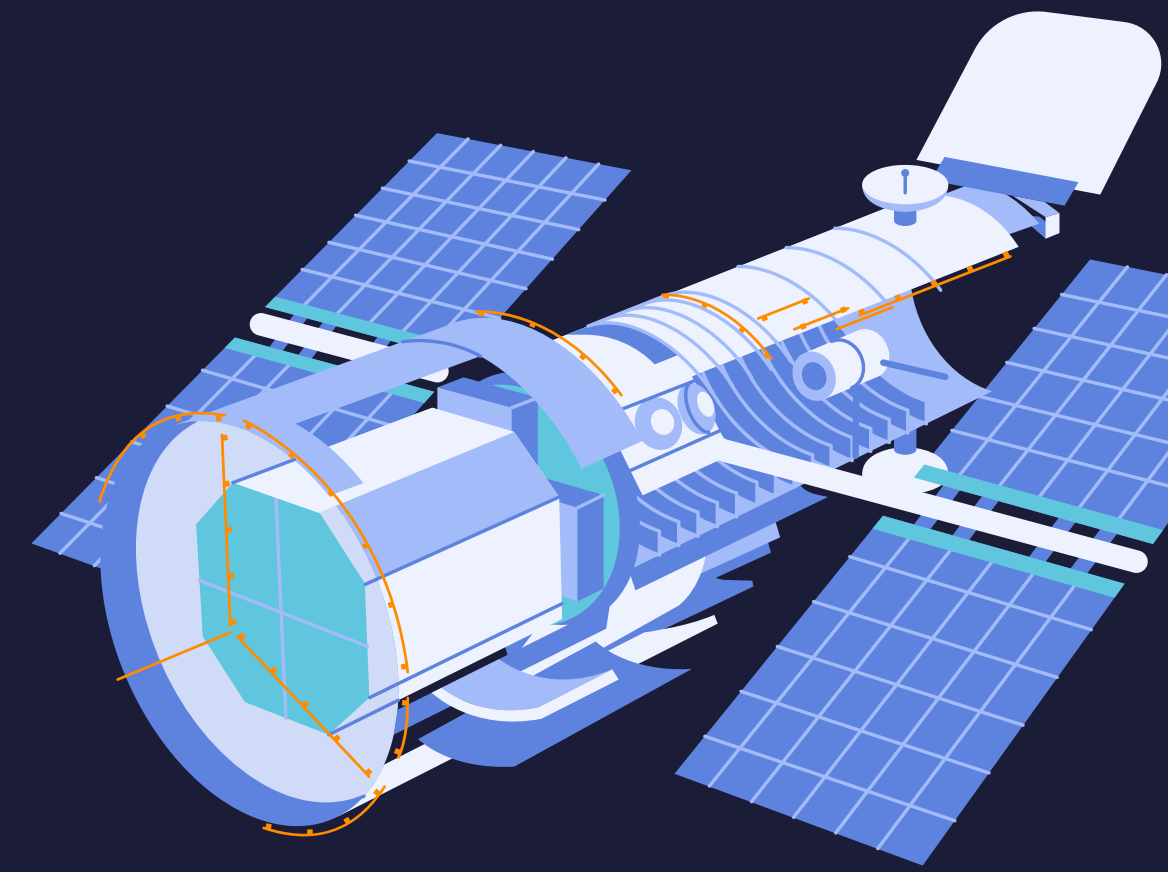- API/Query Language

Documents

Product Overview

```json
{
    "_id": 1,
    "first_name": "Tom",
    "email": "tom@example.com",
    "cell": "765-555-5555",
    "likes": [
        "fashion",
        "spas",
        "shopping"
    ],
    "businesses": [
        {
            "name": "Entertainment 1080",
            "partner": "Jean",
            "status": "Bankrupt",
            "date_founded": {
                "$date": "2012-05-19T04:00:00Z"
            }
        },
        {
            "name": "Swag for Tweens",
            "date_founded": {
                "$date": "2012-11-01T04:00:00Z"
            }
        }
    ]
}
```
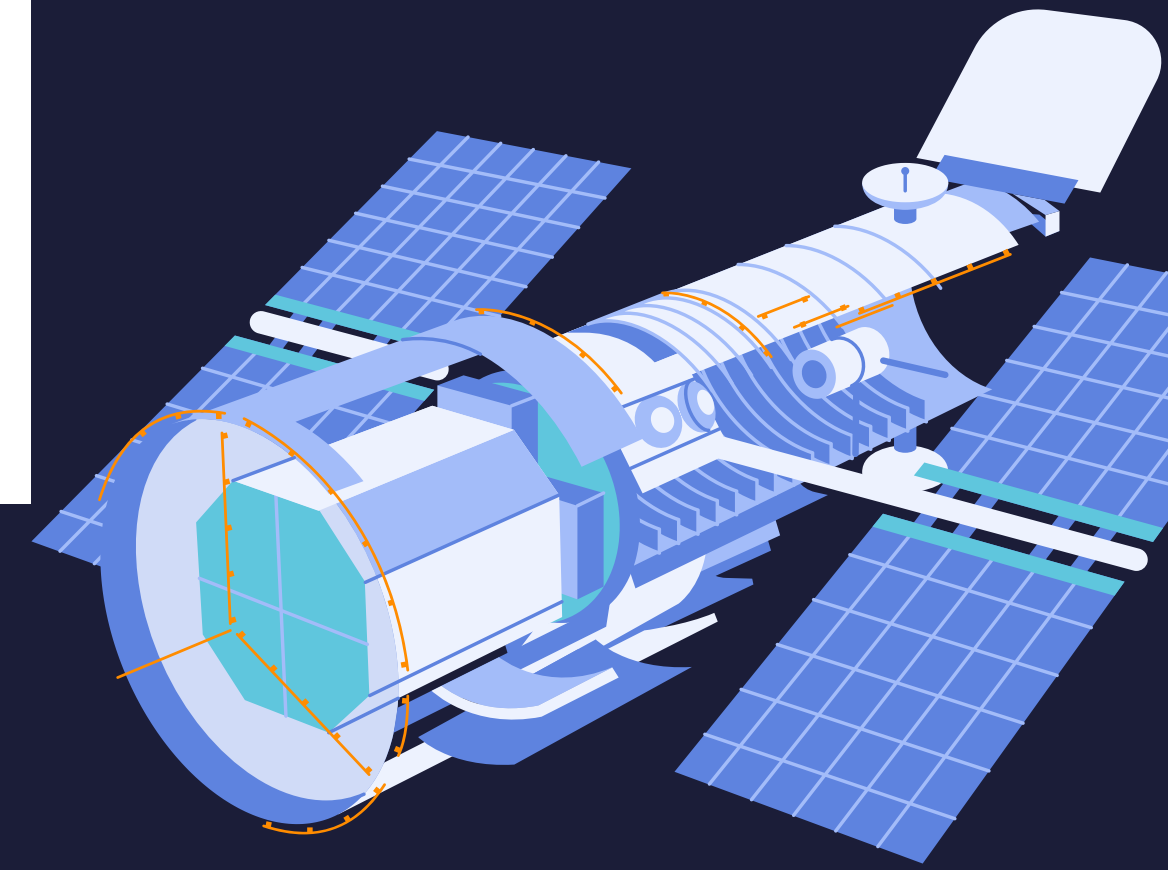
Source: MongoDB

CRUD

Source: MongoDB

```
db.users.insertOne(          ←───── collection
  {
    name: "sue",             ←───── field: value  ⎫
    age: 26,                 ←───── field: value  ⎬ document
    status: "pending"        ←───── field: value  ⎭
  }
)
```
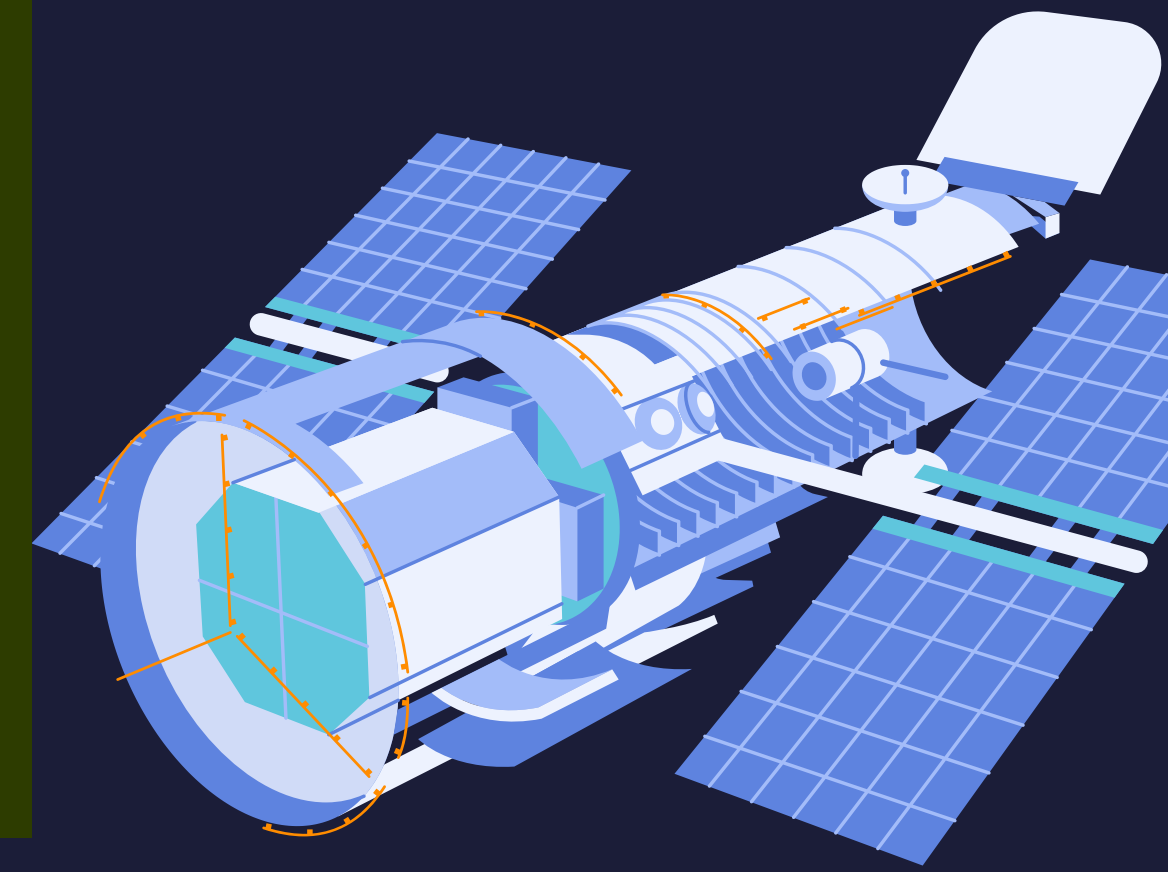
API/Query Language

```
shell> curl http://127.0.0.1:5984
```

Source: CouchDB

```json
{
  "couchdb": "Welcome",
  "version": "3.0.0",
  "git_sha": "83bdcf693",
  "uuid": "56f16e7c93ff4a2dc20eb6acc7000b71",
  "features": [
    "access-ready",
    "partitioned",
    "pluggable-storage-engines",
    "reshard",
    "scheduler"
  ],
  "vendor": {
    "name": "The Apache Software Foundation"
  }
}
```

Feature Comparsions
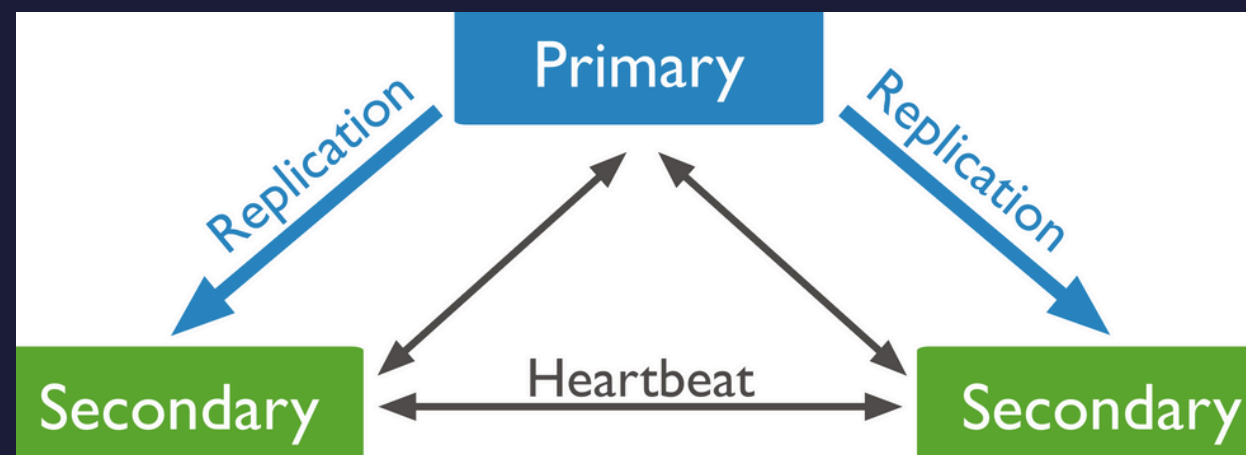
Product Overview

|	CRUD Operations	Flexible Schema	Schema Validation	API/Query Language
MongoDB	Y	Y	Y	Y
CouchDB	Y	Y	N	Y
Firebase	Y	Y	N	Y

MongoDB Architecture

○ Distributed architecture with a "**primary-secondary replication model**" (MongoDB, 2022).

▶ **Replica sets:** multiple copies in different places.

Source: <u>Secondary-Primary Replication Model by MongoDB</u>

○ **Sharding:** Horizontal scaling (data distributed across multiple servers)
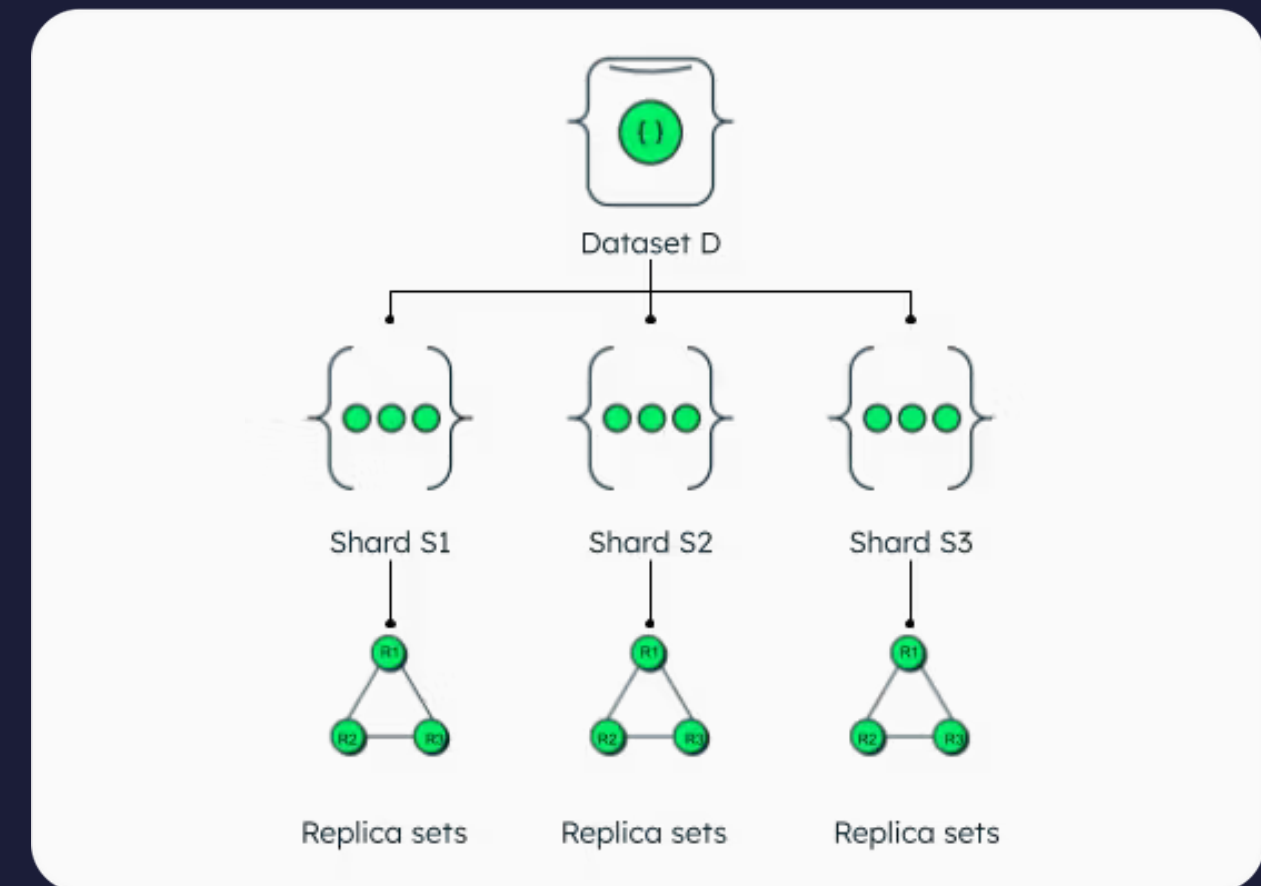
Source: <u>Sharding Model by MongoDB</u>

MongoDB Architecture
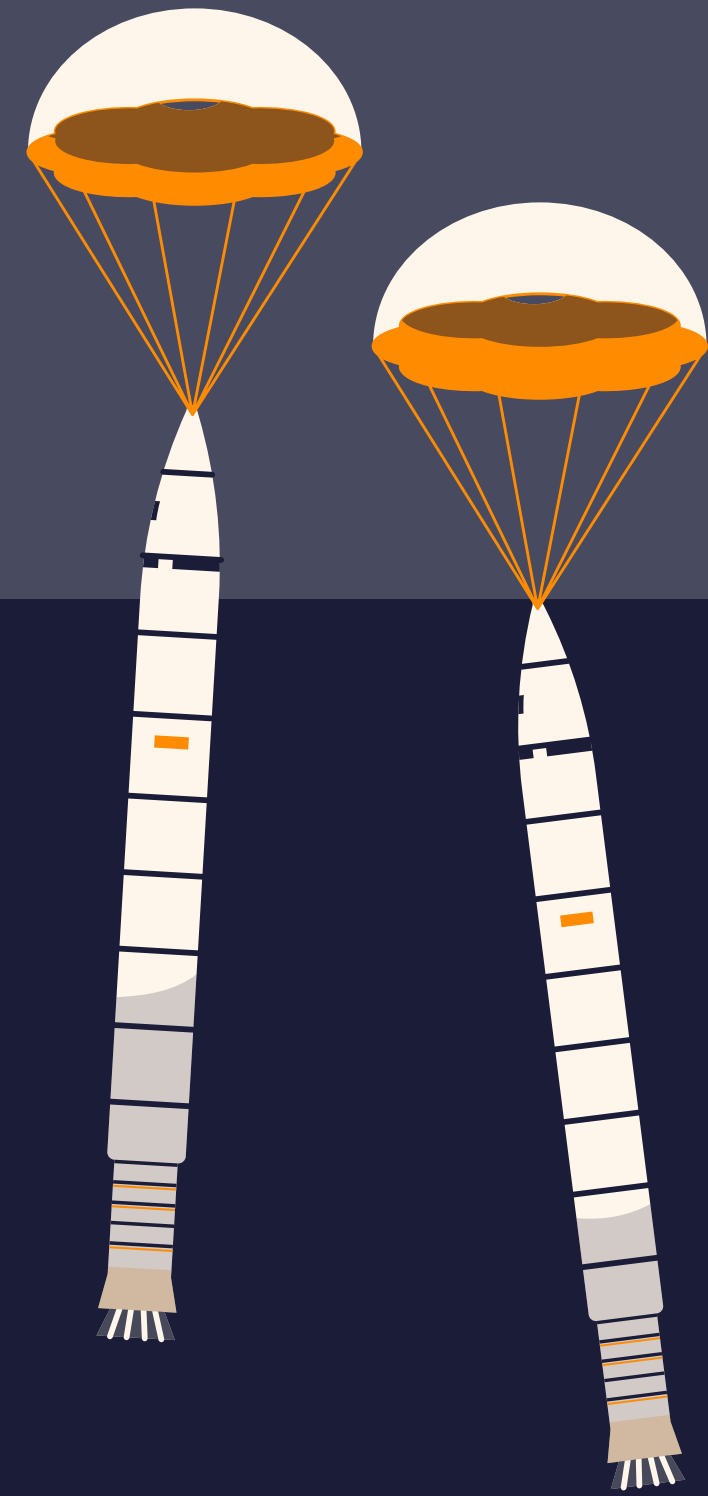
- Data is stored in **BSON** (Binary JSON but more powerful) format
  - BSON adds some "extra" information to documents, like length of strings and subobjects -> faster traversal
  - Fast to encode and decode!

- Document model supports **nested data structures**, making it easy to represent complex relationships.
  - Think of folders that hold all kinds of info, even nested details.

MongoDB
Use Case

- **modern web applications**

- **real-time analytics**

- **IoT:** smart home system

- **cross-platform apps:** travel app (u will find out why)

- **online gaming:** scale to millions of players

- **content management systems:** blogging sites, media libraries

MongoDB

Key Differentiators

- **Flexible Schema:** not too strict data structure, add or change fields anytime!

- **Rich Querying & Indexing:** search & filter in powerful ways

- **Horizontal Scaling:** Spread data across multiple servers

- **High Availability:** Replica sets -> data always available

- **ACID Transactions:** (Atomicity, Consistency, Isolation, and Durability) -> data is safe and consistent even when used by multiple people

- **MongoDB Atlas:** comprehensive cloud service to save the hassle of managing servers/backups

Firebase Architecture

- Runs entirely on **Google Cloud Platform** - Google's global network of servers
  - Highly reliable, fast, scalable
  - Store data across multiple servers in different locations

- **Real-time data sync**
  - Real-time listeners! Data changes in database -> push those changes to all connected devices

- Stores data locally on the device when **offline** -> local save and sync to cloud

- Organize data into **collections** (folders) + **documents** (files)
  - Each document contains key-value pairs (JSON) + nested data

Firebase

Use Case

- **Real-time applications:** chat apps, collaborative tools, gaming leaderboards, and IoT dashboards.
- **Cross-platform mobile app:** offline support and seamless integration with Firebase/Firestore
  - Solid competitor to Supabase and MongoDB

- **Authentication** (very convenient to include Google, Apple, Github Sign in) and other

- **Push Notifications**

Technical Details

Firebase
Key Differentiators

- **Real-time sync:** push updates to all connected devices instantly
- **Offline support:** work without an internet connection + sync data when connected
- **Automatic Scaling & Firestore Integration:** Firestore Auth, Google, Apple sign in
- **Simple & Flexible Data Model:** advanced queries (filter data with multiple conditions)

SIGN UP

Case Study: How Expedia Uses MongoDB

Background Information:
Expedia is one of the world's largest online travel platforms. They utilize **MongoDB** to manage **real-time travel search data**, **customer interaction**, and **booking recommendations.**

How Expedia Leverages MongoDB:
- **Fast Query Performance:** MongoDB stores the travel listings, user preferences, and booking history as a **JSON document**. This allows for the ability to **quickly filter and retrieve data**.
- **Scalability & High Availability:** MongoDB is capable of handing **millions of search queries a day**, and has **99.99% uptime** across all of the **global data centers**. Expedia uses **sharding** to distribute data across **multiple clusters**. This is especially important due to all the spikes in traffic, since the travel industry is highly seasonal.

How Expedia Leverages MongoDB (cont.):

- **Flexible Schemas:** Travel-related data changes **frequently**. Unlike relational databases, MongoDB's schema **flexibility** allows for Expedia to easily incorporate **new data formats** seamlessly **without schema migrations**.
- **Personalized Recommendations:** Expedia uses **MongoDB Atlas**, a cloud service, to store **real-time user behavior**. They analyze and process this data in order to **generate personalized recommendations**, increasing conversion rates.

Conclusion:

Through the use of MongoDB, Expedia has been able to **transform their search experience**. They deliver real time results and personalized recommendations in a **scalable infrastructure** to millions of users. Through the use of MongoDB, they have been able to create a product with **speed and flexibility** enabling them for success.
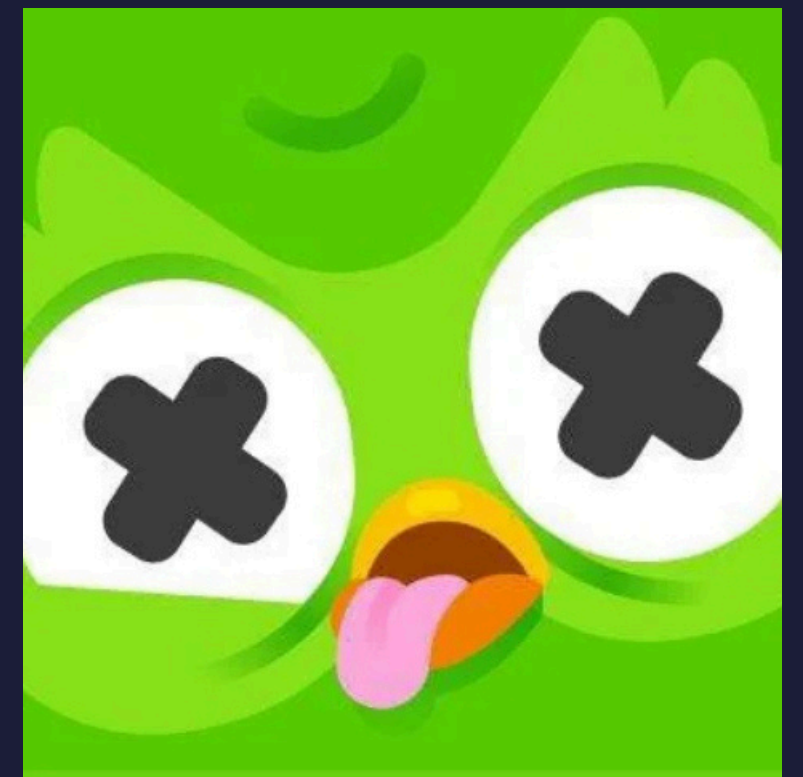
Case Study: How Duolingo Uses Firebase

What is Duolingo: A cross-platform language learning tool where users can complete lessons, progress through a language, and add friends.

Firebase allows Duolingo to have real-time updates, along with providing storage and cloud functions.

How Duolingo leverages Firebase:
- **Realtime Syncing**: Duolingo uses Firebase's realtime syncing of JSON data to provide immediate progress updates while also storing general user info such as languages and friends.
- **Security**: Duolingo must collect user data like age, name, and email along with progress and usage patterns. Therefore, different mechanisms like securing SDKs with Cloud Firestore Security Rules and constraining queries by rules enables Firebase to protect this data.

How Duolingo leverages Firebase (cont.):

- **Scalability**: Duolingo is a fast growing platform with increasing language courses and a rapidly expanding user base. For example, their monthly active users grew 23% from 2021-2022*, and they have over 100 language courses. For this reason, they need a highly scalable database to support this growth trend.
- **Flexibility**: Because Firebase has many integrated products, it is a great option for companies with growing and changing needs.
  *Source: Duolingo

Conclusion: Firebase is a leading document database because of its real time syncing capabilities, scalability, security, and flexibility. All of these features allow companies like Duolingo to give real time updates to their users, while providing scalable and secure data storage.
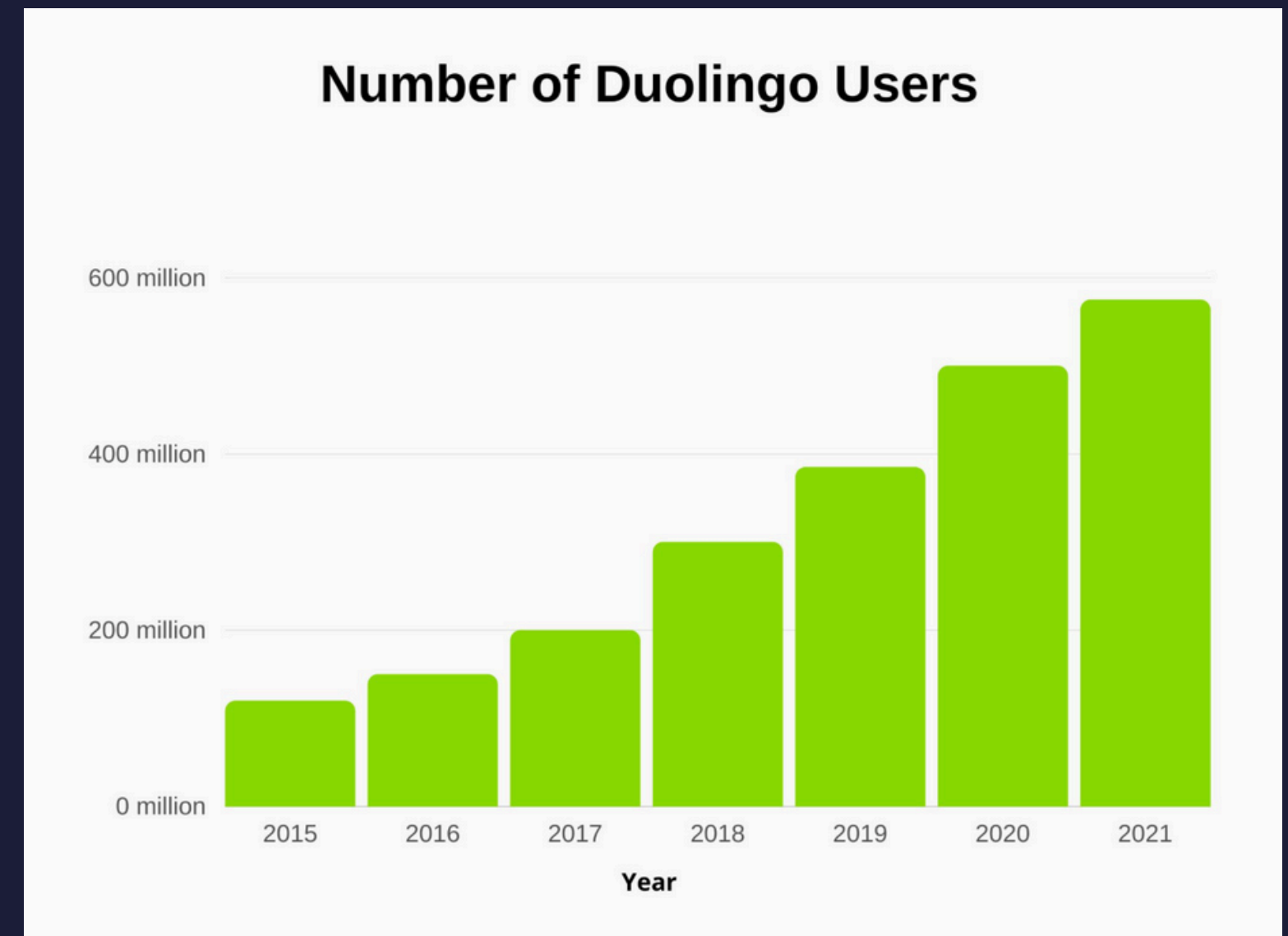

Source: Duolingo Users and Growth Statistics (2024)

NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison

Overview: The increase of Big Data and the limitations of RDBMS over recent years has led to the increasing dominance and significance NoSQL databases, such as document databases. NoSQL databases are better suited for Big Data because they are horizontally scalable, distributed, and more flexible when it comes to modeling increasingly complex and interconnected data.

Contributions: This paper conducts a study on the rise of NoSQL databases, defines and explains the different classifications, and compares these different NoSQL databases to provide readers with a clear understanding of the strengths and weaknesses of NoSQL databases, including document databases.

Relevance: This paper is relevant to document database technology because document databases are a subclass of NoSQL databases, and the authors clearly explain and compare this technology with other NoSQL databases. It also specifically analyzes and compares MongoDB's features to other database products.
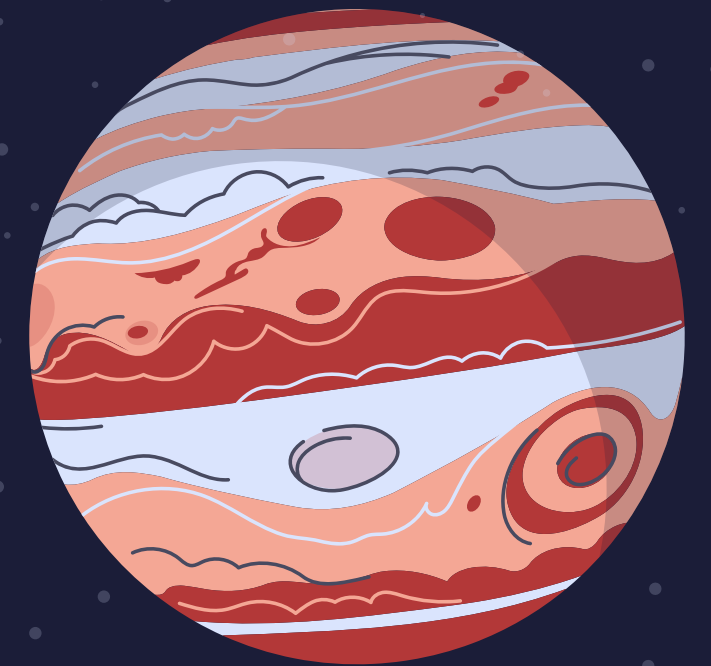
Performance Evaluation of NoSQL Document Databases: Couchbase, CouchDB, and MongoDB

Overview: The study by Carvalho et al. (2023) evaluates the performance of Couchbase, CouchDB, and MongoDB using the Yahoo! Cloud Serving Benchmark (YCSB). Some of the key contributions include:

- A comparative analysis of runtime performance across different workloads and dataset sizes.
- An assessment of scalability when increasing the number of threads and record counts.
- Insights into the strengths and weaknesses of each database for specific operations.

Relevance: This research paper is important for document databases since it looks at the performance of many different document databases. This advances them since it lets developers have more information about when to choose a database based on the use case.
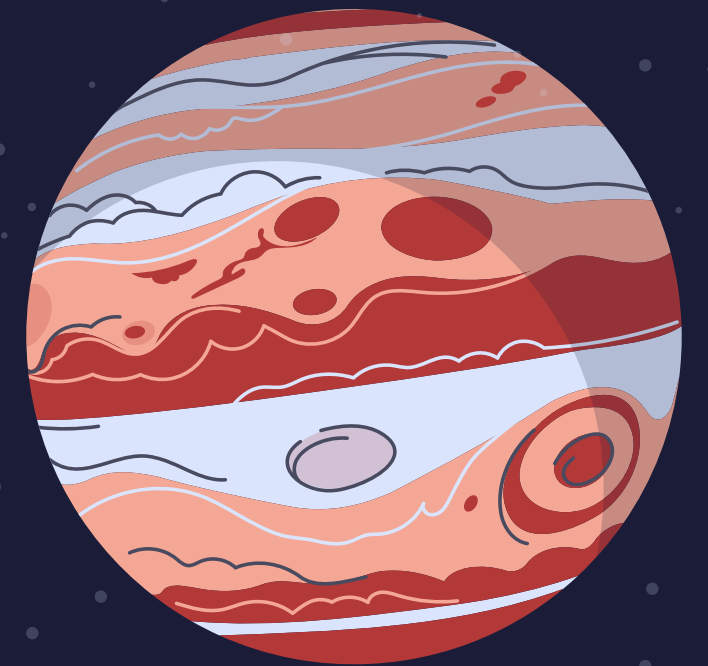
Standing in Markets

Products in the Market

- MongoDB is valued at around $20 billion
- Firebase is a part of the Google ecosystem

Source: https://stockanalysis.com/stocks/mdb/statistics/

Market Analysis

- Gained traction over the last decade and are projected to keep growing over the next 10 years.
- The rise of NoSQL databases has aided the rise in the market of document databases.
  - Driven by the rise in data being generated at higher rates and in more diverse formats.
- North America currently holds the largest share, but Asia is expected to see the most growth in the future.
- Marketing strategy focuses on the flexibility of these systems as well as their ability to process diverse data types.

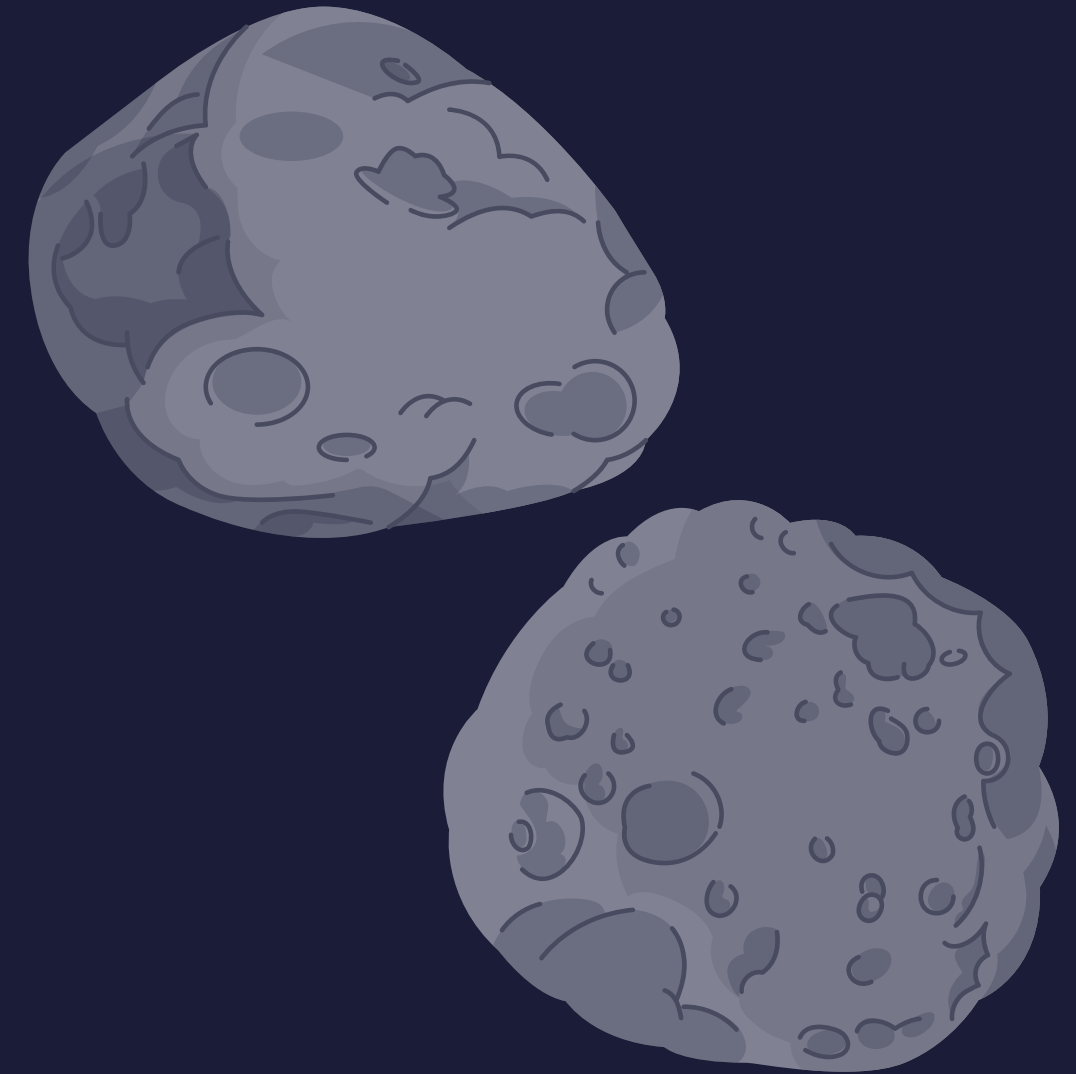Source: https://dataintelo.com/report/document-databases-market
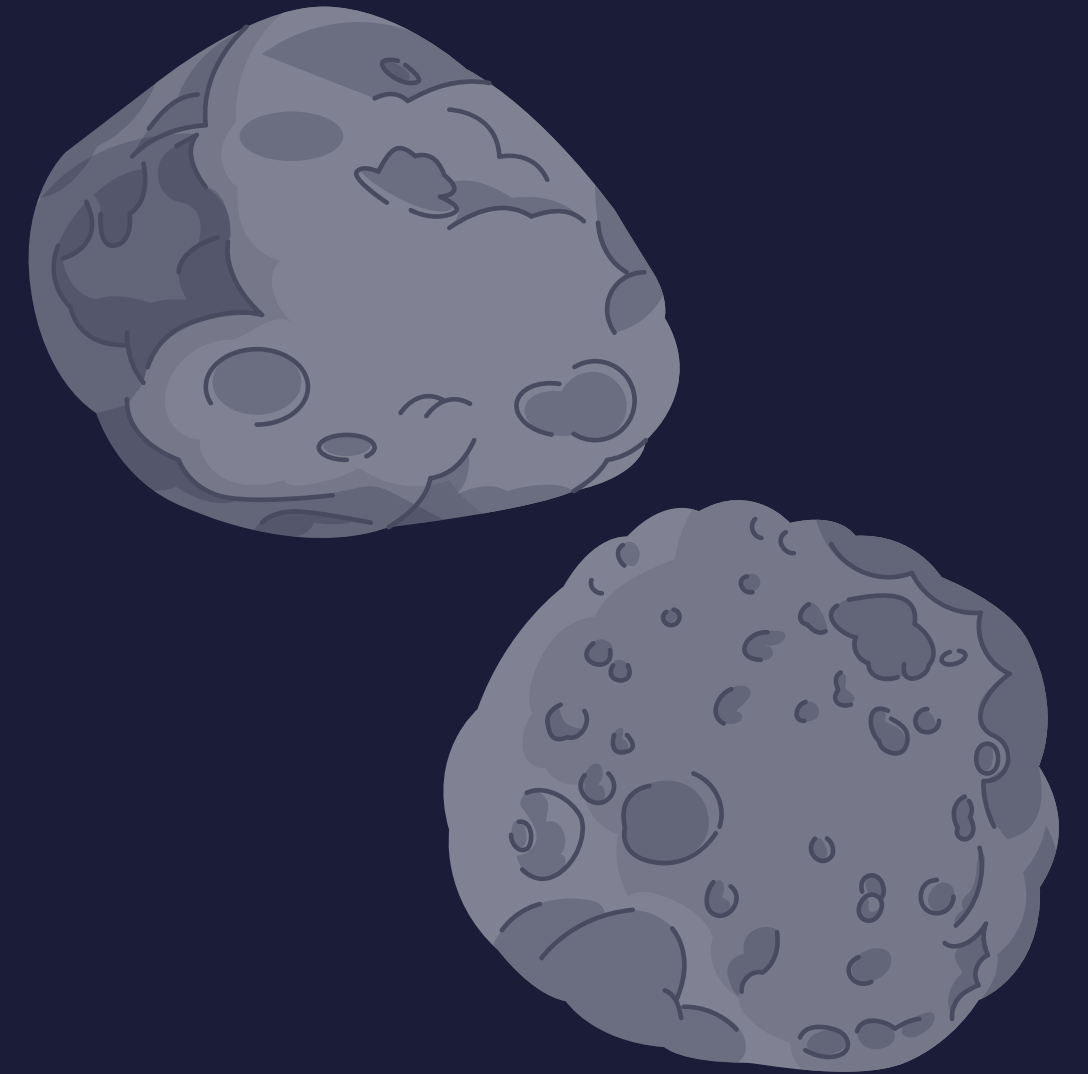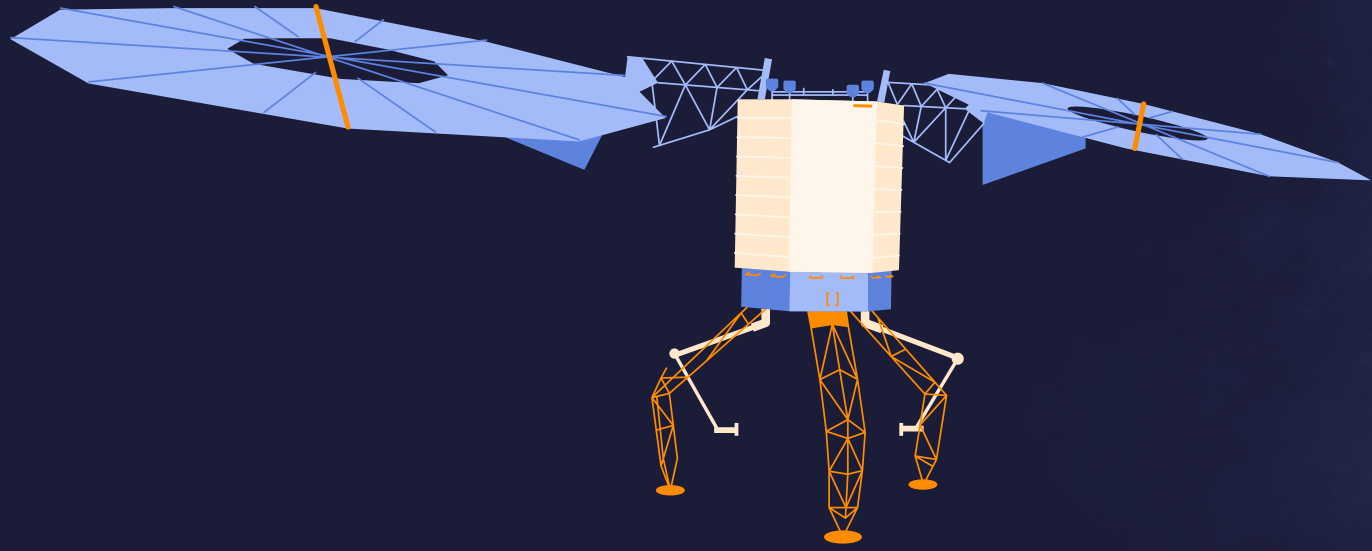
In the Future...

Changes in Tech
- Document Databases will work with AI/ML models more efficiently
  - Data for AI/ML models is often hosted on non-relational databases
- Document Databases will integrate more with data security
  - Rise in data means more security tools are needed for protection

Future Growth
- Document databases are expected to keep growing in usage
  - They are used with AI/ML models, which is a growing field
- They will continue to be used alongside Relational Databases
- Ex. MongoDB
  - MongoDB increased efficiency in 2024 via architectural changes
  - Made it easier to work with encrypted data in 2024
  - Expect further AI integration with their database this year

Sources: https://dataintelo.com/report/document-databases-market, https://www.mongodb.com/blog/post/mongodbs-2024-year-in-review

Any Questions?

Works Cited

- Carvalho, I., Sá, F., & Bernardino, J. (2023). Performance Evaluation of NoSQL Document Databases: Couchbase, CouchDB, and MongoDB. Algorithms, 16(2), 78. https://doi.org/10.3390/a16020078

- Choose a database: Cloud Firestore or realtime database | firebase realtime database. Google. https://firebase.google.com/docs/database/rtdb-vs-firestore

- Duolingo announces record bookings in first quarter 2022 and raises full year guidance. Duolingo, Inc. (2022, May 12). https://investors.duolingo.com/news-releases/news-release-details/duolingo-announces-record-bookings-first-quarter-2022-and-raises

- Duolingo users and Growth Statistics (2024). SignHouse. (2024, August 1). https://usesignhouse.com/blog/duolingo-stats/#interesting-facts-about-duolingo-top-pick-stats

- Moniruzzaman, A. B. M., & Hossain, S. A. (2013). NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison. arXiv [Cs.DB]. Retrieved from http://arxiv.org/abs/1307.0191

- With Expedia Online Travel Gets personal. MongoDB. (n.d.). https://www.mongodb.com/solutions/customer-case-studies/expedia

- Carvalho, I., Sá, F., & Bernardino, J. (2023). Performance Evaluation of NoSQL Document Databases: Couchbase, CouchDB, and MongoDB. Algorithms, 16(2), 78. https://doi.org/10.3390/a16020078

- What is a Document Database?. MongoDB. (n.d.). https://www.mongodb.com/resources/basics/databases/document-databases

- MongoDB CRUD Operations?. MongoDB. (n.d.). https://www.mongodb.com/docs/manual/crud/#std-label-crud

- The Core API. CouchDB. (n.d.). https://docs.couchdb.org/en/stable/intro/api.html

- Stock Analysis. (n.d.). MDB - MongoDB, Inc. Statistics. Retrieved February 24, 2025, from https://stockanalysis.com/stocks/mdb/statistics

- MongoDB. (2024, December 19). MongoDB's 2024 year in review. MongoDB Blog. https://www.mongodb.com/blog/post/mongodbs-2024-year-in-review

- DataIntelo. (n.d.). Document databases market - global industry analysis, size, share, growth, trends, and forecast, 2023-2032. Retrieved February 26, 2025, from https://dataintelo.com/report/document-databases-market

- MongoDB. (2022). *Replication. Mongodb.com.*

- Kesavan, R., Gay, D., Thevessen, D., Shah, J., & Mohan, C. (2023). Firestore: The NoSQL Serverless Database for the Application Developer. *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. https://doi.org/10.1109/icde55515.2023.00259

- *Cloud Firestore*. (n.d.). React Native Firebase. https://rnfirebase.io/firestore/usage

Works Cited Cont.

- Google. (2019). *Cloud Firestore | Firebase*. Firebase. https://firebase.google.com/docs/firestore

- ACID Transactions in MongoDB. (2024, March 12). GeeksforGeeks. https://www.geeksforgeeks.org/acid-transactions-in-mongodb/

- Rathore, M., & Bagui, S. S. (2024). MongoDB: Meeting the Dynamic Needs of Modern Applications. Encyclopedia, 4(4), 1433–1453. https://doi.org/10.3390/encyclopedia4040093

- MongoDB, Inc. (n.d.). Sharding. MongoDB Manual. Retrieved February 25, 2025, from https://www.mongodb.com/docs/manual/sharding/

- MongoDB, Inc. (n.d.). Document databases. MongoDB Resources. Retrieved February 25, 2025