|  |
| --- |
| Module 1: Introduction To MongoDB – Architecture & Installation |

Case Study 2

* **List the advantages and disadvantages of both the plans and select one plan to implement**

SQL vs NoSQL: High-Level Differences

* SQL databases are primarily called as Relational Databases (RDBMS); whereas NoSQL database are primarily called as non-relational or distributed database.
* SQL databases are table based databases whereas NoSQL databases are document based, key-value pairs, graph databases or wide-column stores. This means that SQL databases represent data in form of tables which consists of n number of rows of data whereas NoSQL databases are the collection of key-value pair, documents, graph databases or wide-column stores which do not have standard schema definitions which it needs to adhered to.
* SQL databases have predefined schema whereas NoSQL databases have dynamic schema for unstructured data.
* SQL databases are vertically scalable whereas the NoSQL databases are horizontally scalable. SQL databases are scaled by increasing the horse-power of the hardware. NoSQL databases are scaled by increasing the databases servers in the pool of resources to reduce the load.
* SQL databases uses SQL ( structured query language ) for defining and manipulating the data, which is very powerful. In NoSQL database, queries are focused on collection of documents. Sometimes it is also called as UnQL (Unstructured Query Language). The syntax of using UnQL varies from database to database.
* SQL database examples: MySql, Oracle, Sqlite, Postgres and MS-SQL. NoSQL database examples: MongoDB, BigTable, Redis, RavenDb, Cassandra, Hbase, Neo4j and CouchDb
* For complex queries: SQL databases are good fit for the complex query intensive environment whereas NoSQL databases are not good fit for complex queries. On a high-level, NoSQL don’t have standard interfaces to perform complex queries, and the queries themselves in NoSQL are not as powerful as SQL query language.
* For the type of data to be stored: SQL databases are not best fit for hierarchical data storage. But, NoSQL database fits better for the hierarchical data storage as it follows the key-value pair way of storing data similar to JSON data. NoSQL database are highly preferred for large data set (i.e for big data). Hbase is an example for this purpose.
* For scalability: In most typical situations, SQL databases are vertically scalable. You can manage increasing load by increasing the CPU, RAM, SSD, etc, on a single server. On the other hand, NoSQL databases are horizontally scalable. You can just add few more servers easily in your NoSQL database infrastructure to handle the large traffic.
* For high transactional based application: SQL databases are best fit for heavy duty transactional type applications, as it is more stable and promises the atomicity as well as integrity of the data. While you can use NoSQL for transactions purpose, it is still not comparable and sable enough in high load and for complex transactional applications.
* For support: Excellent support are available for all SQL database from their vendors. There are also lot of independent consultations who can help you with SQL database for a very large scale deployments. For some NoSQL database you still have to rely on community support, and only limited outside experts are available for you to setup and deploy your large scale NoSQL deployments.
* For properties: SQL databases emphasizes on ACID properties ( Atomicity, Consistency, Isolation and Durability) whereas the NoSQL database follows the Brewers CAP theorem ( Consistency, Availability and Partition tolerance )
* For DB types: On a high-level, we can classify SQL databases as either open-source or close-sourced from commercial vendors. NoSQL databases can be classified on the basis of way of storing data as graph databases, key-value store databases, document store databases, column store database and XML databases.

1. MongoDB

[Mongodb](https://www.thegeekstuff.com/2013/01/install-mongodb/) is one of the most popular document based NoSQL database as it stores data in JSON like documents. It is non-relational database with dynamic schema. It has been developed by the founders of DoubleClick, written in C++ and is currently being used by some big companies like The New York Times, Craigslist, MTV Networks. The following are some of MongoDB benefits and strengths:

* Speed: For simple queries, it gives good performance, as all the related data are in single document which eliminates the join operations.
* Scalability: It is horizontally scalable i.e. you can reduce the workload by increasing the number of servers in your resource pool instead of relying on a stand alone resource.
* Manageable: It is easy to use for both developers and administrators. This also gives the ability to shard database
* Dynamic Schema: Its gives you the flexibility to evolve your data schema without modifying the existing data

2. CouchDB

[CouchDB](https://www.thegeekstuff.com/2012/06/install-couch-db/) is also a document based NoSQL database. It stores data in form of JSON documents. The following are some of CouchDB benefits and strengths:

* Schema-less: As a member of NoSQL family, it also have dynamic schema which makes it more flexible, having a form of JSON documents for storing data.
* HTTP query: You can access your database documents using your web browser.
* Conflict Resolution: It has automatic conflict detection which is useful while in a distributed database.
* Easy Replication: Implementing replication is fairly straight forward

3. Redis

[Redis](http://redis.io/) is another Open Source NoSQL database which is mainly used because of its lightening speed. It is written in ANSI C language. The following are some of Redis benefits and strengths:

* Data structures: Redis provides efficient data structures to an extend that it is sometimes called as data structure server. The keys stored in database can be hashes, lists, strings, sorted or unsorted sets.
* Redis as Cache: You can use Redis as a cache by implementing keys with limited time to live to improve the performance.
* Very fast: It is consider as one of the fastest NoSQL server as it works with the in-memory dataset.
* **Choose a data model after analyzing the source files provided to you with this case study**

Since the format is in

TV Show contains ->

1. Show details

2. Seasons (Many)

3. Episodes (Many)

4. Reviews (Many)

5. Cast and Crew (Many)

and also it can be in text, video, audio, graphics – MongoDB is preferred to do this design

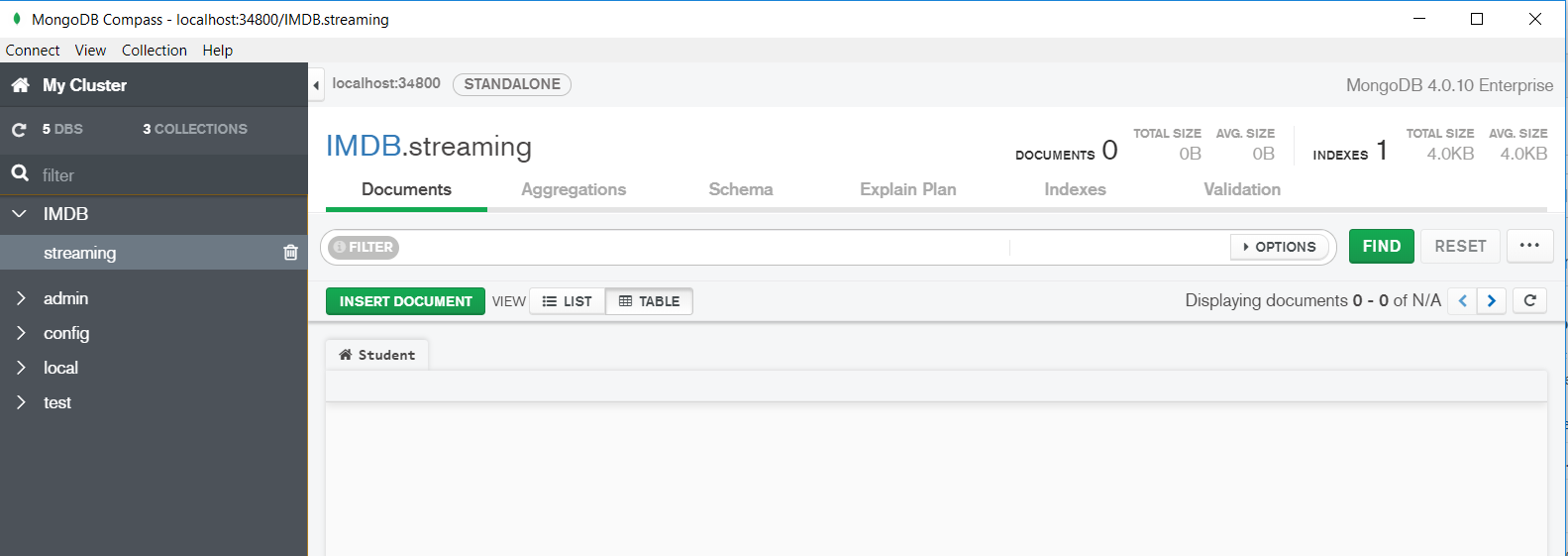
* **Select appropriate data model and create your personalized server**

MongoDB is an open-source, nonrelational database that provides support for JSON-like, document-oriented storage systems. It supports a flexible data model that enables you to store data of any structure, and provides a rich set of features, including full index support, sharding, and replication. AWS enables you to set up the infrastructure to support MongoDB deployment in a flexible, scalable, and cost-effective manner in the AWS Cloud.

Database with name ‘IMDB’ with port number 34800 have been created

* **Create collections in your database**

‘Streaming’ collection for gathering the informations



* **Insert records in each collection**

