Hussien Tarek

Part 1:



3. Cassandra and MongoDB: ACID vs. BASE Transaction Management Comparison Briefing

1.1 Cassandra

ACID Transaction Management:

- Supports atomicity on row & partition level but not on multiple records
 operations level so, conflict in operations success on different replicas or
 subset of records processes won't result in instantaneous rollback.
- Provides flexible eventual consistency as it has two types & one of each is tunable: -

- Tunable Consistency (in terms of CAP, most up-to-date transaction):
 Allow managing consistency & availability trade-offs by defining replicas policy to acknowledge a read or write operation successfully.
- 2. Linearizable Consistency (in terms of ACID, transaction doesn't corrupt database): Allow simulation of some degree of atomicity & serial isolation level for transactions by evaluating responses of different replica nodes to read or write operations & prioritize the latest timestamp response in case of inconsistency using a "compare-and-set" (CAS) request.
- Earlier versions of Cassandra allowed other users visibility of partial row update but now it allows full-row level isolation across different users such that partial updates are only visible for those users who performed the update operation.
- Provide data durability globally by replicating it across multiple nodes & locally by saving data in memory & through a commit log.

BASE Transaction Management:

- Cassandra is primarily designed to prioritize high availability. This is achieved through its distributed storage system, which allows for linear scalability across multiple data centers or within the same node.
- To ensure this high availability, Cassandra employs eventual consistency as a
 compromise, although it does offer flexibility in other areas as mentioned
 before. For lightweight transactions that require strong consistency,
 Cassandra provides linearizable consistency. Additionally, users can define
 the number of replicas required to reach consensus when reading updates
 from the database, allowing for tunable consistency.
- Has a soft state due to the eventual consistency.

1.2 MongoDB

ACID Transaction Management:

- Supports atomicity at a single document level and its subdocuments.
 MongoDB 4.2 versions support atomicity on multiple document level with a performance impact [3].
- By default, MongoDB offers strong consistency. The primary servers in a
 MongoDB cluster are highly consistent compared to the secondary servers
 which only provide eventual consistency. Consequently, accessing data from
 secondary nodes is not permitted by default However, this can be modified to
 prioritize availability over consistency.
- Allow for document update full isolation & rolling back the update in case of an error & provide isolation patterns & operators like "update if current" that prevent updates if the document changed after the last read & "\$isolation" that ensures update operation isolated from other operations.
- Uses the same traditional database durability model that provides tuning of durability & performance trade-offs through configuring transactions committing to the database after saving updates to the journal files. by parameters like "syncdelay" & "journalCommitInterval".

Based on the information available, it can be inferred that earlier versions of MongoDB only supported ACID compliance at the document level. However, starting from version 4.0, MongoDB introduced support for multi-document ACID transactions. Furthermore, in version 4.2, distributed multi-document ACID transactions were also implemented; although this came with a trade-off in terms of performance [4].

BASE Transaction Management:

- MongoDB ensures availability by implementing replica sets in each data center. However, it prioritizes consistency over availability.
- As previously stated, MongoDB ensures strong consistency of primary server data and by default limits access to secondary servers for consistent reads.
- Has a soft state due to the eventual consistency of secondary MongoDB servers.

1.3 Conclusion

The transaction management capabilities of Cassandra and MongoDB showcase their distinctive design principles and priorities. Cassandra places emphasis on Basic Availability, Soft State, and Eventual Consistency rather than strong consistency. On the other hand, MongoDB offers comprehensive support for ACID transactions along with robust guarantees of strong consistency. Determining the most suitable database for your application depends on specific requirements and trade-offs that need to be considered between availability, partition tolerance, and consistency.

	Atomicity	Consistency	Isolation	Durability
Cassandra	Row-level	Eventual	Row-level	Durable
MongoDB	Single document	Eventual	Isolation present	Durable

Table 1 ACID Properties on Cassandra & MongoDB [5].

	Basically Available	Eventually Consistent
Cassandra	Highly available	Eventual consistency
	Example: Distributed Storage.	Example: Tunable for partition
	Multi data center. Linear	tolerance. Linearized for strong
	scalability.	consistency.
MongoDB	Consistency over availability	Strong consistency
	Example: Uses replica sets.	Example: Consistent data only on
	Distributed across datacenters.	primary MongoDB server. Reads
		allowed only from primary.

Table 2 BASE Properties on Cassandra & MongoDB [5].

		-

Part 2:

1) MongoDB Lab

Setup:

Set an account on MongoDB Atlas - https://cloud.mongodb.com

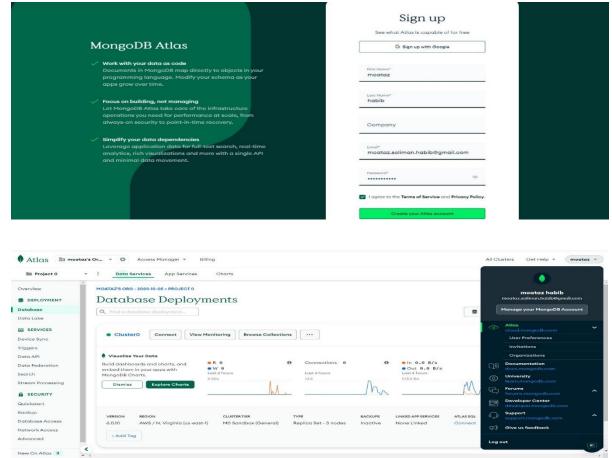


Figure 2 MongoDB Setup

Load the Sample Netflix Movies Database to your Data Lake

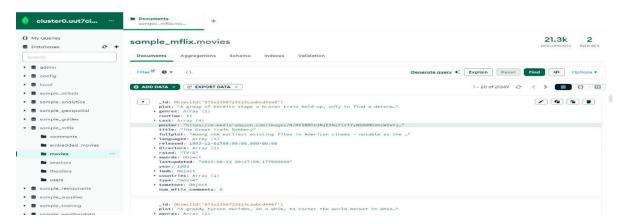


Figure 3 Load Data

 Set up a connection to this database instance from MongoDB compass or any other MongoDB client.

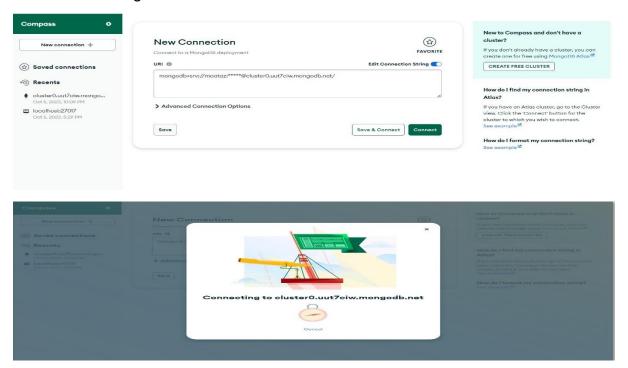


Figure 4 Connection Setup

1. Briefly describe the movies database document model.

The movies database document model consists of a collection named "movies" that contains documents representing movies.

Each movie document has various fields.

The document model allows for flexible and dynamic data storage, as each movie document can have different fields and values.

Field	Description	Count	Categorization of Documents
_id	The unique identifier for this document in MongoDB.	21349	Required field
plot	A brief synopsis of the movie.	20203	Optional field
genres	An array listing the genres relevant to the movie.	21237	Optional field
runtime	The runtime of the movie in minutes.	20910	Optional field
rated	The rating for the movie. (e.g., "TV-G","R", "G", TV-MA").	11455	Optional field

cast	An array listing the actors featured in the movie.	20987	Optional field
num_mflix_comments	The number of comments made in the MFLIX system.	21349	Optional field
poster	A URL to the poster image of the movie.	18044	Optional field
title	The title of the movie.	21349	Required field
fullplot	A full synopsis of the movie plot.	19852	Optional field
languages	An array of languages in which the movie is available.	21119	Embedded fields
released	The release date of the movie.	20878	Optional field
directors	An array listing the directors of the movie.	21107	Optional field
writers	An array listing the writers of the movie.	20256	Optional field
awards	An object detailing the awards the movie has won.	21349	Embedded field
lastupdated	The timestamp of the last update made to this document.	21349	Optional field
year	The year the movie was released.	21349	Optional field
imdb	An object with details about the film's IMDB rating, votes, and ID.	21349	Embedded field
countries	An array of countries where the movie was released.	21339	Embedded field
type	Defines the type of the document, in this case, a "movie".	21349	Optional field
tomatoes	An object with ratings from Rotten Tomato reviews.	18588	Embedded field
metacritic	Metacritic score or rating for the movie.	6964	embedded field

2. Filter the documents for type "movies" that are released before 1970 and rated as "PASSED."

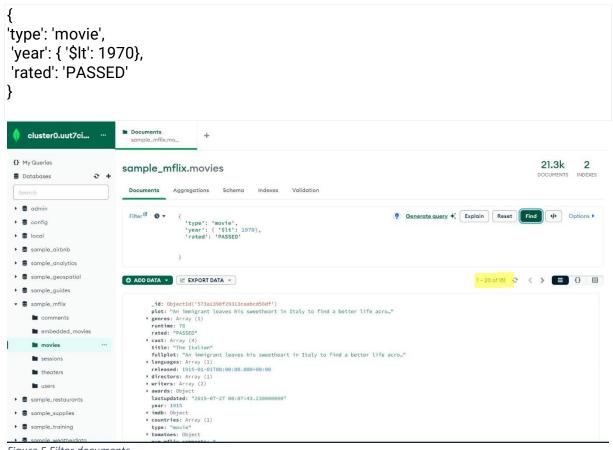


Figure 5 Filter documents

3. Build an Aggregation Pipeline that shows all entries of type movie that have won at least one award and return the release year aggregate counts.

```
[
{
$match: {
    $and: [
       { type: "movie" },
       { awards: { $exists: true, $ne: [] } },
       { "awards.wins": { $gte: 1 } }
    1
  }
},
 $group: {
  _id: "$year",
  count: {
  $sum: 1
   }
  }
},
   $sort: {
  id: 1
                                    Aggregations
sample_mflix.mo..
   cluster0.uut7ci...
 {} My Queries
                                                                                                     21.3k
                     sample_mflix.movies
 Databases
 • 🛢 admin
                                                                       (4) (E)
                                type: "movie" },
awards: { $exists: true, $ne: [] } },
"awards.wins": { $gte: 1 } }
 ▼ 3 sample_mflix
```

Figure 6 Aggregation Pipeline

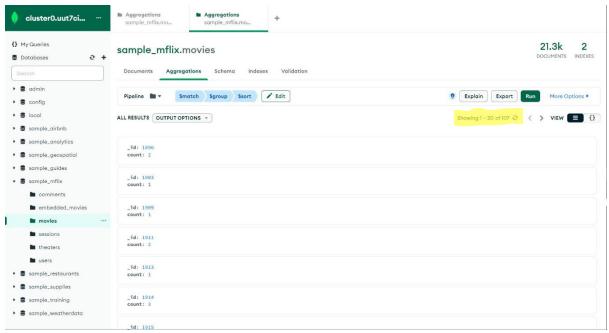
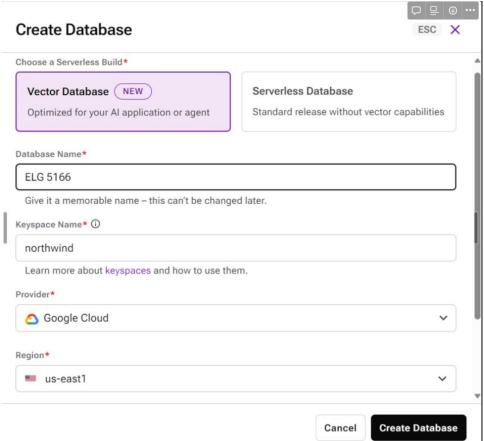


Figure 7 Aggregation Pipeline con.

2) Cassandra Lab:

Figure 8 Cassandra Setup



- Create a Keyspace called northwind
- Create the customer tables

Figure 9 customer tables

Load the attached data:

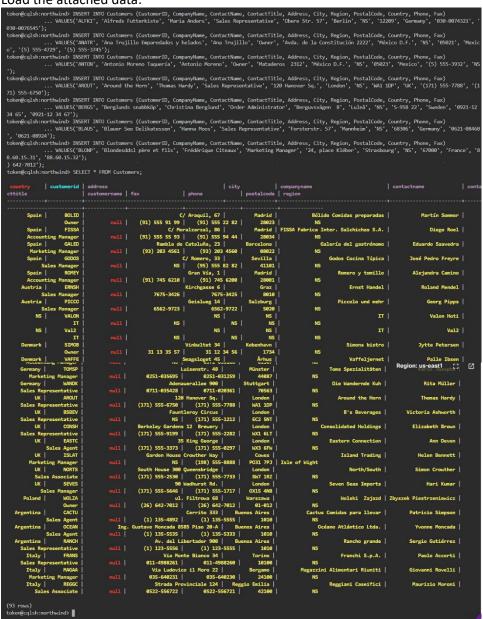


Figure 10 Insertion statements

Queries:

1. Provide the query and the results (screenshots and a copy of your query) that show the customers from Rio de Janeiro, Brazil ordered by their addresses.

```
The answer:
```

```
CREATE TABLE CustomersByCountryCity (
Country TEXT,
City TEXT,
CustomerID TEXT,
CompanyName TEXT,
ContactName TEXT,
ContactTitle TEXT,
Address TEXT,
Region TEXT,
PostalCode TEXT,
Phone TEXT,
Fax TEXT,
PRIMARY KEY ((Country, City), Address, CustomerID)
);
```

The query:

SELECT *

FROM CustomersByCountryCity

WHERE Country = 'Brazil' AND City = 'Rio de Janeiro';

```
CREATE TABLE northwind.customersbycountrycity (
     country text,
     address text,
     customerid text,
     companyname text,
     contactname text,
     contacttitle text,
     fax text,
    phone text,
  region text,
PRIMARY KEY ((country, city), address, customerid)
WITH CLUSTERING ORDER BY (address ASC, customerid ASC)
AND additional_write_policy = '99p'
    AND bloom_filter_fp_chance = 0.01
AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
AND comment = ''
     AND compaction = {'class': 'org.apache.cassandra.db.compaction.UnifiedCompactionStrategy'}
AND compression = {'chunk_length_in_kb': '16', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
     AND crc_check_chance = 1.0
     AND default_time_to_live = 0
     AND gc_grace_seconds = 864000
     AND max_index_interval = 2048
     AND memtable flush period in ms = 0
     AND min_index_interval = 128
     AND read_repair = 'BLOCKING'
     AND speculative_retry = '99p';
```

Figure 11 Table_1 creation

Figure 12 first query

2. Provide a list of customers that are in the Sales Manager role without forcing the scan of all partitions across all databases. The result should be ordered by their names

```
The answer:
CREATE TABLE CustomersByContactTitle (
 ContactTitle TEXT,
 CustomerID TEXT,
 ContactName TEXT,
 Country TEXT,
 City TEXT,
 CompanyName TEXT,
 Address TEXT,
 Region TEXT,
 PostalCode TEXT,
 Phone TEXT,
 Fax TEXT,
 PRIMARY KEY (ContactTitle, ContactName, CustomerID)
);
The query:
SELECT * FROM CustomersByContactTitle
WHERE ContactTitle = 'Sales Manager'
ORDER BY ContactName;
```

Figure 13 Table_ creation



Figure 14 Second query

List of Tables

Table 1 ACID Properties on Cassandra & MongoDB [5]
Table 2 BASE Properties on Cassandra & MongoDB [5]
Table 3 Table demonstrate movies database

