

## Cloud Analytics

### Assignment 1

<b>Student name:</b> Ahmed Ibraheem Ahmed Badwy	<b>Student ID:</b> 300389393
<b>Student name:</b> Anas Ibrahim Ali Elbattr	<b>Student ID:</b> 300389368
<b>Student name:</b> Esmael Alahmad yEbrahim Ezz	<b>Student ID:</b> 300389885
<b>Student name:</b> Yousef Abdelhaleem Ahmed Shindy	<b>Student ID:</b> 300389365

University of Ottawa, Ca

---

By typing in my name and student ID on this form and submitting it electronically, I am attesting to the fact that I have reviewed not only my work but the work of my team member, in its entirety.

I attest to the fact that my work in this project adheres to the fraud policies as outlined in the Academic Regulations in the University's Graduate Studies Calendar. I further attest that I have knowledge of and have respected the "Beware of Plagiarism" brochure for the university. To the best of my knowledge, I also believe that each of my group colleagues has also met the aforementioned requirements and regulations. I understand that if my group assignment is submitted without a completed copy of this Personal Work Statement from each group member, it will be interpreted by the school that the missing student(s) name is confirmation of the nonparticipation of the aforementioned student(s) in the required work.

We, by typing in our names and student IDs on this form and submitting it electronically,

- warrant that the work submitted herein is our own group members' work and not the work of others
- acknowledge that we have read and understood the University Regulations on Academic Misconduct
- acknowledge that it is a breach of University Regulations to give or receive unauthorized and/or unacknowledged assistance on a graded piece of work

## Part 1

Q1 –

- **NoSQL databases, or "not only SQL,":**

are designed for handling large volumes of diverse, unstructured data with flexible models, including document, key-value, wide-column, and graph databases.

- **Neo4j:**

is a type of graph database that is a versatile database capable of storing and querying complex data relationships, facilitating real-time analytics like fraud detection and product recommendations, and enabling the development of intelligent applications that learn and reason over big data, such as social networks and fraud detection systems.

Q2 –

- **Carbon Footprint Analysis:**

The problem of climate change poses a significant global challenge, with the urgent need to reduce greenhouse gas emissions and mitigate its adverse impacts. Big data plays a crucial role in addressing this problem by offering a comprehensive and data-driven approach to understanding, monitoring, and combating climate change. Through the collection and analysis of vast amounts of environmental, energy, and emissions data, big data empowers researchers, organizations, and policymakers to make informed decisions and develop effective strategies for reducing carbon footprints. It enables real-time monitoring of energy consumption, identifies emissions sources, and supports the optimization of transportation and energy usage. Furthermore, big data facilitates predictive modeling, allowing for better preparation and adaptation to the changing climate patterns. Its contribution extends to improving renewable energy integration, fostering sustainable practices in agriculture, and enhancing climate communication efforts to raise awareness.

- **Challenges:**

However, while big data offers immense potential in addressing climate change, it also presents several challenges. One major hurdle is data quality and standardization, as information from various sources may vary in accuracy and consistency. Privacy concerns also arise when collecting and sharing personal or sensitive data related to energy consumption and transportation. Additionally, the sheer volume of data can overwhelm existing infrastructure and necessitate significant computational resources. Interpretation of big data results and translating them into actionable policies can be complex, requiring skilled data scientists and analysts. Finally, the cost of implementing and maintaining big data systems may be prohibitive for smaller organizations or developing nations. Overcoming these challenges requires a coordinated effort among stakeholders to establish data standards, address privacy issues, invest in infrastructure, and foster data literacy to harness the full potential of big data in the fight against climate change.

### Q3 -

- **Cassandra:**

**BASE:**

Cassandra follows a BASE (Basically Available, Soft state, eventually consistent) model. It prioritizes high availability and partition tolerance over strict consistency.

**ACID:**

Cassandra does not provide full ACID transactions like traditional relational databases. Instead, it offers tunable consistency levels, allowing you to choose the level of consistency for each operation, which can range from strong consistency to eventual consistency.

- **MongoDB:**

**BASE:**

MongoDB traditionally follows a more flexible, BASE-like model, but it also supports ACID transactions in recent versions.

**ACID**

ACID transactions in MongoDB provide the ability to perform multi-document transactions with properties like atomicity, consistency, isolation, and durability, similar to traditional relational databases.

### Q4 -

- **The potential NOSQL databases:**

Apache Kafka, Amazon DynamoDB, InfluxDB, Elasticsearch and Redis

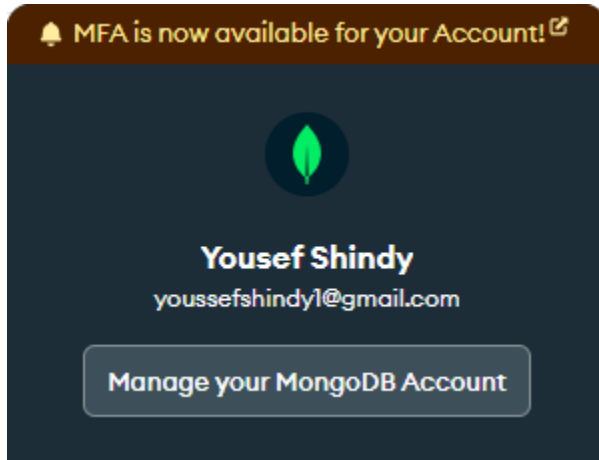
- **DynamoDB:**

is an ideal choice for capturing and archiving data from IoT devices due to its scalability, performance, high availability, and cost-effectiveness. It scales horizontally to handle IoT-generated data, performs well with millions of reads and writes, and ensures data collection even during device downtime. Design considerations include choosing appropriate partition and sort keys, provisioning throughput based on expected traffic, and setting data retention with TTL values for automatic deletion when needed.

## Part 2 NoSQL Labs

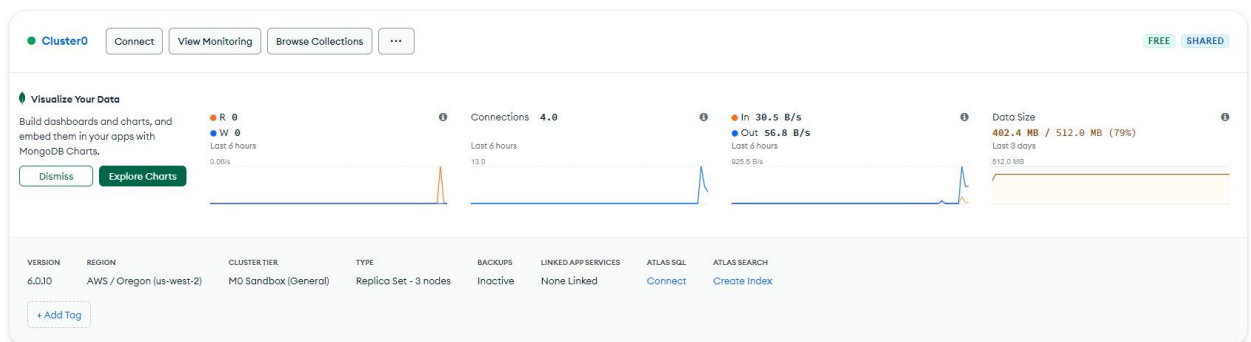
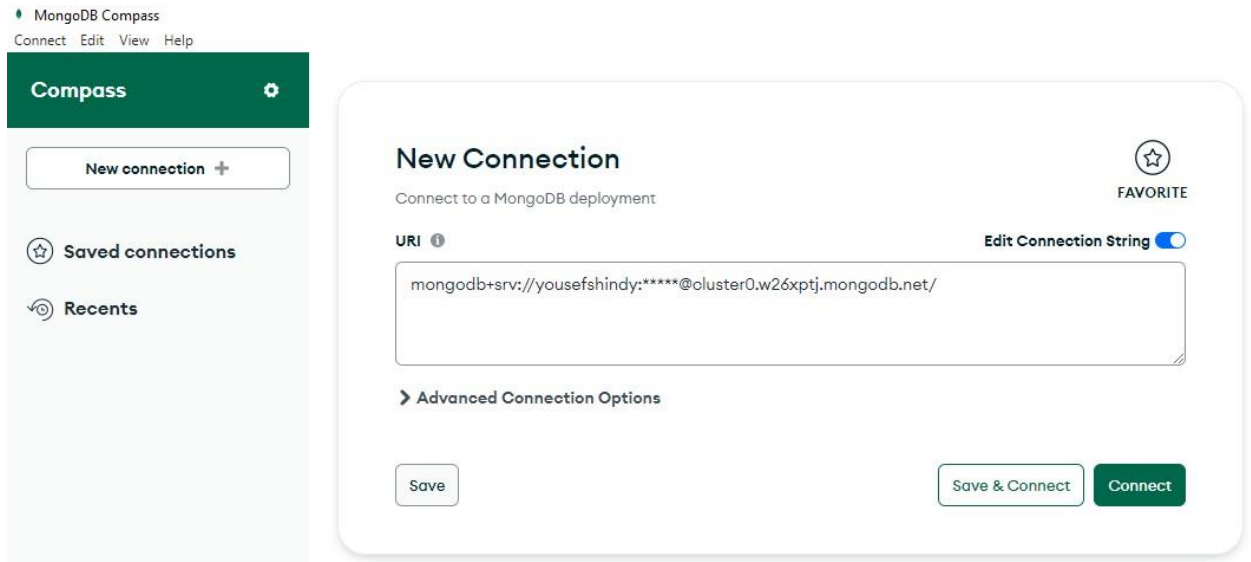
### 1) MongoDB Lab

- Set an account on MongoDB Atlas



- Set up a connection to this database instance from MongoDB compass

We copy the connection string to connect to the cluster from MongoDB and paste it into compass after changing the password to establish the connection.



Then we open the sample\_mflix dataset and select the movies.

The screenshot shows the MongoDB Atlas web interface. On the left, a sidebar lists various databases and collections. The 'sample\_mflix' database is expanded, and the 'movies' collection is selected. The main panel displays the 'sample\_mflix.movies' collection. At the top, there are tabs for 'Documents', 'Aggregations', 'Schema', 'Indexes', and 'Validation'. Below these tabs, there is a 'Filter' button and a text input for a query. Below the input, there are 'ADD DATA' and 'EXPORT DATA' buttons. The main area shows a JSON document for a movie. The document includes fields like '\_id', 'plot', 'genres', 'runtime', 'cast', 'poster', 'title', 'fullplot', 'languages', 'released', 'directors', 'rated', 'awards', 'lastupdated', 'year', 'imdb', 'countries', 'type', 'tomatoes', and 'num\_mflix\_comments'.

```
{
  "_id": "573a1390f29313caabcd42e8",
  "plot": "A group of bandits stage a brazen train hold-up, only to find a determ...",
  "genres": [
    "Action",
    "Crime"
  ],
  "runtime": 11,
  "cast": [
    "John Ford",
    "John Wayne",
    "Vivian Vance",
    "John Ford"
  ],
  "poster": "https://m.media-amazon.com/images/H/WV5BMTU3NjE5NzYtYTYyNS00MDVmLWlWYj...",
  "title": "The Great Train Robbery",
  "fullplot": "Among the earliest existing films in American cinema - notable as the ...",
  "languages": [
    "English"
  ],
  "released": "1903-12-01T00:00:00.000+00:00",
  "directors": [
    "John Ford"
  ],
  "rated": "TV-G",
  "awards": {
    "nominations": 0,
    "wins": 0
  },
  "lastupdated": "2015-08-13 00:27:59.177000000",
  "year": 1903,
  "imdb": {
    "id": "tt0012828",
    "rating": 8.5,
    "votes": 1000000
  },
  "countries": [
    "USA"
  ],
  "type": "movie",
  "tomatoes": {
    "score": 0,
    "comments": 0
  },
  "num_mflix_comments": 0
}
```

Provide the following answers:

1- Briefly describe the movies database document model

The movies database contains 22 field some of which have important information like:

- **\_id**: A unique identifier for the document
- **awards**: Contains information about the number of nominations and wins.
- **year**: The year in which the movie was released.
- **imdb**: An array that contains the rating, number of votes, and the imdb id.
- **tomatoes**: Contains information about the rotten tomatoes rating and comments for the movies.
- **poster**: A URL pointing to the movie's poster image.
- **type**: The type of media such as movie or series.
- **writers**: An array containing the writers of the media.
- **runtime**: The duration of the movie or series in minutes.
- **fullplot**: A more detailed description of the plot of the movie.
- **countries**: An array the contains the list of countries associated with the movie.
- **cast**: An array that contains the cast members such as directors or actors.
- **languages**: An array of the languages that the movies is available in.
- **metacritic**: Contains the score of the movie on metacritic.
- **num\_mflix\_comments**: The number of comments related to the movie in mflix.
- **lastupdated**: A timestamp of the last time the document was updated.
- **released**: The release date of the movie in ISO 8601 format.
- **title**: The title of the movie.
- **rated**: The movie's content rating.
- **genres**: An array that lists the genre of the movie.
- **directors**: An array that lists the directors of the movie.
- **plot**: Contains a brief description of the plot of the movie.

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

```
{
  'type': 'movie',
  'year': {'$lt': 1970},
  'rated': 'PASSED'
}
```

The screenshot shows a MongoDB query interface. At the top, a filter query is entered: `{ 'type': 'movie', 'year': {'$lt': 1970}, 'rated': 'PASSED' }`. Below the query, there are buttons for "Generate query", "Explain", "Reset", and "Find". The results section shows two documents. The first document is for a movie titled "The Italian" (1915), which is a silent film. The second document is for a movie titled "At 10 years old, Genna becomes a ragged orphan when his sainted mother..." (1913), which is a silent film.

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: {
      'type': 'movie',
      'awards.wins': { $gte: 1 }
    }
  },
  {
    $group: {
      _id: '$year',
      count: { $sum: 1 }
    }
  }
]
```

The screenshot shows a MongoDB aggregation pipeline interface. On the left, the pipeline is defined as follows:

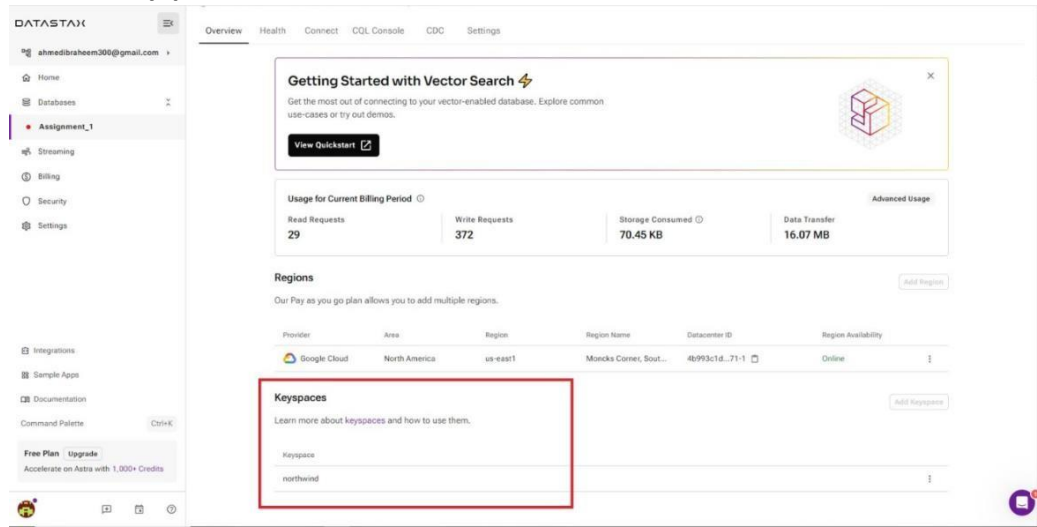
```
1 [
2   {
3     $match: {
4       'type': 'movie',
5       'awards.wins': { $gte: 1 }
6     },
7   },
8   {
9     $group: {
10      _id: '$year',
11      count: { $sum: 1 }
12    }
13  }
14 ]
```

On the right, the "PIPELINE OUTPUT" section shows a sample of 10 documents. The output is a list of years and their corresponding counts:

_id	count
1945	21
1953	41
1922	3
1913	1
1989	158
2000	433

## 2) Cassandra Lab

- Create a Keyspace called northwind



- Create two tables

### Table One:

CREATE TABLE IF NOT EXISTS Customers (

CustomerID TEXT,

CompanyName TEXT,

ContactName TEXT,

ContactTitle TEXT,

Address TEXT,

City TEXT,

Region TEXT,

PostalCode TEXT,

Country TEXT,

Phone TEXT,

Fax TEXT,

PRIMARY KEY ((City, Country), Address, CustomerID)

);

**Table Two:**

```
CREATE TABLE IF NOT EXISTS Customers_role(

    CustomerID TEXT ,

    CompanyName TEXT,

    ContactName TEXT,

    ContactTitle TEXT,

    Address TEXT,

    City TEXT,

    Region TEXT,

    PostalCode TEXT,

    Country TEXT,

    Phone TEXT,

    Fax TEXT,

    PRIMARY KEY ((ContactTitle ),ContactName,CustomerID)

);
```

- Load the attached data into your tables

city	country	address	customerid	companyname	contactname	contacttitle	fax	phone	postalcode	region
Helsinki	Finland	Keskuskatu 45	WILMC	Wilman Kala	Matti Karttunen	Owner/Marketing Assl.				us-east1
Portland	USA	89 Chiaroscuro Rd.	LONEP	Lonesome Pine Restaurant	Fran Wilson	Sales Manager		(503) 555-9040		
Portland	USA	89 Jefferson Way Suite 2	THEBI	The Big Cheese	Liz Nixon	Marketing Manager				
Berlin	Germany	Obere Str. 57	ALFKI	Alfreds Futterkiste	Maria Anders	Sales Representative		030-0076545		
Cornwall	Germany	Taucherstraße 10	QUICK	QUICK-Stop	Horst Kloss	Accounting Manager				
Toulouse	France	1 rue Alsace-Lorraine	LAPAI	La maison d'Asie	Annette Roulet	Sales Manager		61.77.61.11		
Charleroi	Belgium	Boulevard Tirou, 255	SUPRD	Suprêmes délices	Pascale Cartrain	Accounting Manager		(071) 23 67 22 21		
Lisboa	Portugal	Estrada da saúde n. 58	PRIME	Princesa Isabel Vinhos	Isabel de Castro	Sales Representative				
Lisboa	Portugal	Jardim das rosas n. 32	FURIB	Furia Bacalhau e Frutos do Mar	Lino Rodriguez	Sales Manager		(1) 354-2535		
Madrid	Spain	C/ Araquil, 67	BOLID	Bóldo Conidas preparadas	Martin Sommer	Owner		(91) 555 91 99		
Madrid	Spain	C/ Moralzarzal, 86	FISSA	FISSA Fabrica Inter. Salchichas S.A.	Diego Roel	Accounting Manager		(91) 555 55 93		
Madrid	Spain	Gran Vía, 1	ROMEY	Romero y tomillo	Alejandra Camino	Accounting Manager		(91) 745 6210		
Münster	Germany	Luisenstr. 48	TOUSP	Toms Spezialitäten	Karin Josephs	Marketing Manager		0251-035695		
Lander	USA	P.O. Box 555	SPLIR	Split Rail Beer & Ale	Art Braunschweiger	Sales Manager		(307) 555-6525		
Tübingen	Germany	City Center Plaza 516 Main St.	HUNGC	Hungry Coyote Import Store	Yoshi Latimer	Sales Representative		(503) 555-2370		
Brandenburg	Germany	Muehlstr. 90	KOENE	Königlich Essen	Philip Cramer	Sales Associate				
Bremen	Germany	59 rue de l'Abbaye	VINET	Vins et alcools Chevalier	Paul Henriot	Accounting Manager		26.47.15.11		
Comex	UK	Garden House Crowther Way	ISLAT	Island Trading	Helen Bennett	Marketing Manager				

## Queries:

- show the customers from Rio de Janeiro, Brazil ordered by their addresses

```
SELECT * FROM customers
```

```
WHERE city = 'Rio de Janeiro' AND country = 'Brazil'
```

```
ORDER BY address;
```

```
token@cqlsh:northwind> SELECT * FROM customers
... WHERE city = 'Rio de Janeiro' AND country = 'Brazil'
... ORDER BY address;
```

city	country	address	customerid	companyname	contactname	contacttitle	fax	phone	postalcode	region
Rio de Janeiro	Brazil	Av. Copacabana, 267	RICAR	Ricardo Adocicados	Janete Limeira	Assistant Sales Agent		(21) 555-3412	02389-890	RJ
Rio de Janeiro	Brazil	Rua da Panificadora, 12	QUEDE	Que Delícia	Bernardo Batista	Accounting Manager		(21) 555-4252	02389-673	RJ
Rio de Janeiro	Brazil	Rua do Paço, 67	HANAR	Hanari Carnes	Mario Pontes	Accounting Manager		(21) 555-8765	05454-876	RJ

```
(3 rows)
token@cqlsh:northwind>
```



- Provide a list of customers that are in the Sales Manager role ordered by their names

```
SELECT * FROM Customers_role
WHERE ContactTitle = 'Sales Manager'
ORDER BY ContactName;
```

```
token@cqlsh:northwind> SELECT * FROM Customers_role
... WHERE ContactTitle = 'Sales Manager'
... ORDER BY ContactName;
```

contacttitle	contactname	customerid	address	city	companyname	country	fax	phone	postalcode	region
Sales Manager	Annette Roulet	LAMAI	1 rue Alsace-Lorraine	Toulouse	La maison d'Asie	France	61-77-61-11	61-77-61-10	31000	NS
Sales Manager	Art Braunschweiler	SPLIR	P.O. Box 555	Lander	Split Rail Beer & Ale	USA	(307) 555-6525	(307) 555-4680	82520	WY
Sales Manager	Fran Wilson	LONEP	89 Chiaroscuro Rd.	Portland	Lonesome Pine Restaurant	USA	(503) 555-9646	(503) 555-9573	97219	OR
Sales Manager	Georg Pippes	PICCO	Geislweg 14	Salzburg	Piccolo und mehr	Austria	6562-9723	6562-9722	5020	NS
Sales Manager	Hari Kumar	SEVES	90 Wadhurst Rd.	London	Seven Seas Imports	UK	(171) 555-5646	(171) 555-1717	OX15 4NB	NS
Sales Manager	José Pedro Freyre	GODOS	C/ Romero, 33	Sevilla	Godos Cocina Típica	Spain	NS	(95) 555 82 82	41101	NS
Sales Manager	Lino Rodriguez	FURIB	Jardim das rosas n. 32	Lisboa	Furia Bacalhau e Frutos do Mar	Portugal	(1) 354-2535	(1) 354-2534	1675	NS
Sales Manager	Michael Holz	RICSU	Grenzacherweg 237	Genève	Richter Supermarkt	Switzerland	NS	0897-034214	1203	NS
Sales Manager	Palle Ibsen	VAFFE	Smagsloget 45	Århus	Vaffeljernet	Denmark	86 22 33 44	86 21 32 43	8200	NS
Sales Manager	Paula Parente	MELLI	Rua do Mercado, 12	Resende	Wellington Importadora	Brazil	NS	(14) 555-8122	08737-363	SP
Sales Manager	Roland Mendel	ERNSH	Kirchgasse 6	Graz	Ernst Handel	Austria	7675-3426	7675-3425	8010	NS

(11 rows)

## References:

- [1] <https://neo4j.com/docs/>
- [2] Jordan, Gregory. Practical Neo4j. Apress, <https://doi.org/10.1007/978-1-4842-0022-3>.
- [3] Papadopoulos, Thanos, and ME Balta. "Climate Change and Big Data Analytics: Challenges and Opportunities." International Journal of Information Management, vol. 63, 2022, <https://doi.org/10.1016/j.ijinfomgt.2021.102448>.
- [4] <https://cassandra.apache.org/doc/latest/>
- [5] Hewitt, Eben. Cassandra : the Definitive Guide. O'Reilly, 2011.
- [6] <https://www.mongodb.com/docs/>
- [7] Chodorow, Kristina. MongoDB : the Definitive Guide. Second edition., O'Reilly, 2013.
- [8] <https://docs.aws.amazon.com/dynamodb/>