

Universidade Federal da Paraíba – Campus I
Centro de Informática
Departamento de Informática

Big Data: conceitos e aplicações
Laboratório 4: Modelo de Dados Colunar

SETUP DO AMBIENTE

Aluna: Emmanuella Faustino Albuquerque

- 1) Criar uma conta e banco de dados no DataStax/AstraDB.

<https://astra.datastax.com/>

https://docs.datastax.com/en/astra-serverless/docs/manage/db/managing-db.html#create_your_astra_db_database

- 2) Instalar CQLSQL shell e configurar para conectar no banco de dados criado anteriormente.

https://docs.datastax.com/en/astra-serverless/docs/connect/cql/connect-cqlsh.html#standalone_cql_shell

```
Bundle:
Databases > Connect > Select a Method: Drivers (Legacy) >
Download Bundle

$ chmod +x ./bin/cqlsh
$ cd /cqlsh-astra
$ ./bin/cqlsh -u **clientID** -p **clientSecret** -b
/path/to/secure-connect**database_name**.zip
```

EXERCÍCIOS

- 1) Crie um script (arquivo de texto comum) com uma sequência de instruções CQL trabalhando com o banco de dados Cassandra.

Os arquivos encontram-se na pasta src/scripts.cql e src/CRUD.cql

```
cqlsh >
CREATE TABLE example.chats_by_theme (
    chat_id uuid,
    theme text,
    created_at timestamp,
    owner_id uuid,
    description text,
    owner_from text,

    -- partitioning, ordering,
    PRIMARY KEY ((theme), created_at)
) WITH CLUSTERING ORDER BY (created_at DESC);

-- Create

INSERT INTO example.chats_by_theme (
    chat_id,
    theme,
    created_at,
    owner_id,
    description,
    owner_from
) VALUES (
    UUID(),
    'Discussion',
    toTimeStamp(NOW()),
    UUID(),
    'BLACK VS GREEN',
    'EUA'
);

INSERT INTO example.chats_by_theme (
    chat_id,
    theme,
    created_at,
```

```

    owner_id,
    description,
    owner_from
) VALUES (
    UUID(),
    'Blurt Out',
    toTimeStamp(NOW()),
    UUID(),
    'rosas são vermelhas violetas são azuis',
    'BR'
);

INSERT INTO example.chats_by_theme (
    chat_id,
    theme,
    created_at,
    owner_id,
    description,
    owner_from
) VALUES (
    UUID(),
    'Discussion',
    toTimeStamp(NOW()),
    UUID(),
    'ice cream or açaí',
    'EUA'
);

-- only test (not run in production)
SELECT * FROM example.chats_by_theme;

```

Saída:

theme	created_at	chat_id	description	owner_from	owner_id
Blurt Out	2022-10-27 01:07:28.107000+0000	c12c042b-d5f8-4410-8374-9a9261b2ae23	rosas são vermelhas violetas são azuis	BR	ef458d9f-c0e0-4270-b5a6-b1057380d4a4
Discussion	2022-10-27 01:07:42.265000+0000	be25c740-4179-4803-9a34-63c931229f80	ice cream or açaí	EUA	29bec205-3bd7-4b38-ada8-579a72caafe7
Discussion	2022-10-27 01:07:21.496000+0000	b8b66326-b40b-4026-a8b3-f6ec1b6249e1	BLACK VS GREEN	EUA	83532107-c72c-4634-bbbf-ed7ba29f6a0a

- 2) Definir um esquema para 2 tabelas para entidades de tipos diferentes
 - Defina pelo menos uma coluna para cada um dos seguintes tipos de dados: tupla, lista, conjunto e mapa;
 - Insira cerca de 5 linhas em cada uma de suas tabelas

```

cqlsh >
-- #####
-- 2.

-- Tupla (Tuple), Lista (List)

CREATE TABLE example.client (
  client_id uuid PRIMARY KEY,
  client_name text,
  address tuple<varchar, varchar, varchar>, -- país, estado,
cidade
  balance list<float>
);

-- Conjunto (Set), Mapa (Map)

CREATE TABLE example.topic (
  topic_id uuid PRIMARY KEY,
  topic_name text,
  tags set<text>, -- unique values
  shares map<text, int>
);

-- CLIENT ENTITY
INSERT INTO example.client ( client_id, client_name, address,
balance )
VALUES (
  UUID(),
  'Ana',
  ('Brasil', 'Paraíba', 'João Pessoa'),
  [110.50, 500.70, 900.55]
);

INSERT INTO example.client ( client_id, client_name, address,
balance )
VALUES (
  UUID(),
  'Guilherme',
  ('Brasil', 'Paraíba', 'Cajazeiras'),
  [400, 200, 1000]
);

```

```
INSERT INTO example.client ( client_id, client_name, address,
balance )
VALUES (
    UUID(),
    'Marta',
    ('Brasil', 'Paraíba', 'Sousa'),
    [80.50, 600, 170]
);
```

```
INSERT INTO example.client ( client_id, client_name, address,
balance )
VALUES (
    UUID(),
    'Antônio',
    ('Brasil', 'Paraíba', 'Campina Grande'),
    [10.50, 900, 30]
);
```

```
INSERT INTO example.client ( client_id, client_name, address,
balance )
VALUES (
    UUID(),
    'Letícia',
    ('Brasil', 'Paraíba', 'Patos'),
    [209, 800.70, 10]
);
```

-- TOPIC ENTITY

```
INSERT INTO example.topic ( topic_id, topic_name, tags, shares )
VALUES (
    UUID(),
    'dead man found in mansion',
    { 'tragedy', 'news' },
    { 'facebook': 70, 'instagram': 100 }
);
```

```
INSERT INTO example.topic ( topic_id, topic_name, tags, shares )
VALUES (
    UUID(),
    'mendes live attracts millions',
    { 'music', 'live' },

```

```

        { 'facebook': 1000, 'instagram': 2000 }
    );

INSERT INTO example.topic ( topic_id, topic_name, tags, shares )
VALUES (
    UUID(),
    'fundraising for the lost cat',
    { 'money', 'help' },
    { 'facebook': 90, 'instagram': 10 }
);

INSERT INTO example.topic ( topic_id, topic_name, tags, shares )
VALUES (
    UUID(),
    'new movie about multiverses draws attention',
    { 'cinema', 'movie' },
    { 'facebook': 1000, 'instagram': 5000 }
);

INSERT INTO example.topic ( topic_id, topic_name, tags, shares )
VALUES (
    UUID(),
    'the queen lost her son and father',
    { 'news' },
    { 'facebook': 5, 'instagram': 20 }
);

```

```
SELECT * FROM example.client;
```

Saída:

client_id	address	balance	client_name
f88a4056-0589-4cea-85f5-32e2d29ccb4b	('Brasil', 'Paraíba', 'Sousa')	[80.5, 600, 170]	Marta
c60e0907-c8e4-49a0-8ea8-4c4287d48d31	('Brasil', 'Paraíba', 'Cajazeiras')	[400, 200, 1000]	Guilherme
e7bfafd8-786d-4b88-9849-32dff8576cf	('Brasil', 'Paraíba', 'João Pessoa')	[110.5, 500.70001, 900.54999]	Ana
5305df9b-4478-4192-a8c3-b78be8fbd702	('Brasil', 'Paraíba', 'Campina Grande')	[10.5, 900, 30]	Antônio
4e9009a4-06b3-45c4-84e9-c4304cf01fd8	('Brasil', 'Paraíba', 'Patos')	[209, 800.70001, 10]	Letícia

```
SELECT * FROM example.topic;
```

Saída:

topic_id	shares	tags	topic_name
dde345c3-ab0c-4c00-9905-4dfd55cc448e	{'facebook': 90, 'instagram': 10}	{'help', 'money'}	fundraising for the lost cat
5380db97-da57-4e06-b35a-44aae5372cc7	{'facebook': 1000, 'instagram': 5000}	{'cinema', 'movie'}	new movie about multiverses draws attention
ee14ca8a-ee22-47e5-aa3c-b446f892f638	{'facebook': 1000, 'instagram': 2000}	{'live', 'music'}	mendes live attracts millions
459f789e-f264-4d75-958b-91c6d71ff6a2	{'facebook': 70, 'instagram': 100}	{'news', 'tragedy'}	dead man found in mansion
6b334761-4778-4e4e-8662-bdcd3330e938	{'facebook': 5, 'instagram': 20}	{'news'}	the queen lost her son and father

3) Escreva pelo menos 3 declarações de atualização

- Execute operações de substituição, adição e remoção de primitivas (todas) em colunas de todos os tipos de coleção (todas);
- Ou seja, você deve envolver pelo menos 9 operações primitivas diferentes em tais colunas

```
cqlsh >
-- Update

-- #####
-- Tuple is frozen (address)

SELECT * FROM example.client WHERE client_id =
5305df9b-4478-4192-a8c3-b78be8fbd702;

-- substituição
UPDATE example.client SET address = ('Brasil', 'Paraíba',
'Pombal') WHERE client_id = 5305df9b-4478-4192-a8c3-b78be8fbd702;

--
#####
-- List (balance)

SELECT * FROM example.client WHERE client_id =
5305df9b-4478-4192-a8c3-b78be8fbd702;

-- adição
UPDATE example.client SET balance = balance + [200] WHERE
client_id = 5305df9b-4478-4192-a8c3-b78be8fbd702;

--
#####
-- Set (tags)

SELECT * FROM example.topic WHERE topic_id =
6b334761-4778-4e4e-8662-bdcd3330e938;

-- adição
UPDATE example.topic SET tags = tags + {'lost'} WHERE topic_id =
6b334761-4778-4e4e-8662-bdcd3330e938;

-- remoção
```

```

UPDATE example.topic SET tags = tags - {'lost'} WHERE topic_id =
6b334761-4778-4e4e-8662-bdcd3330e938;

--
#####
-- Map (shares)

SELECT * FROM example.topic WHERE topic_id =
6b334761-4778-4e4e-8662-bdcd3330e938;

-- substituição
UPDATE example.topic SET shares = shares + {'facebook': 1000}
WHERE topic_id = 6b334761-4778-4e4e-8662-bdcd3330e938;

UPDATE example.topic SET shares['facebook'] = 200 WHERE topic_id =
6b334761-4778-4e4e-8662-bdcd3330e938;

-- adição
UPDATE example.topic SET shares = shares + {'youtube': 30} WHERE
topic_id = 6b334761-4778-4e4e-8662-bdcd3330e938;

-- remoção
UPDATE example.topic SET shares = shares - {'youtube'} WHERE
topic_id = 6b334761-4778-4e4e-8662-bdcd3330e938;

--
#####

-- Delete

-- Client

SELECT * FROM example.client;

DELETE FROM example.client WHERE client_id =
4e9009a4-06b3-45c4-84e9-c4304cf01fd8;

-- Topic

SELECT * FROM example.topic;

DELETE FROM example.topic WHERE topic_id =

```



```
dde345c3-ab0c-4c00-9905-4dfd55cc448e;
```

Algumas Alterações:

```
SELECT * FROM example.topic;
```

topic_id	shares	tags	topic_name
5380db97-da57-4e06-b35a-44aa5372cc7	{'facebook': 1000, 'instagram': 5000}	{'cinema', 'movie'}	new movie about multiverses draws attention
ee14ca8a-ee22-47e5-aa3c-b446f892f638	{'facebook': 1000, 'instagram': 2000}	{'live', 'music'}	mendes live attracts millions
459f789e-f264-4d75-958b-91c6d71ff6a2	{'facebook': 70, 'instagram': 100}	{'news', 'tragedy'}	dead man found in mansion
6b334761-4778-4e4e-8662-bdcd3330e938	{'facebook': 200, 'instagram': 20, 'youtube': 30}	{'lost', 'news'}	the queen lost her son and father

4) Escreva 3 instruções select

- Use as cláusulas WHERE e ORDER BY pelo menos uma vez (ambas);
- Use ALLOW FILTERING em uma consulta que não pode ser avaliada sem esta instrução

Referência: <http://cassandra.apache.org/doc/latest/cql/>

```
cqlsh >
-- #####
-- 4.

-- WHERE e ORDER BY

SELECT * FROM example.client WHERE client_id =
f88a4056-0589-4cea-85f5-32e2d29ccb4b;

SELECT * FROM example.chats_by_theme WHERE theme = 'Discussion'
ORDER BY created_at DESC;

-- ALLOW FILTERING

SELECT * FROM example.chats_by_theme WHERE owner_from = 'BR' ALLOW
FILTERING;
```

WHERE

```
SELECT * FROM example.chats_by_theme WHERE theme = 'Discussion'
ORDER BY created_at;
```

Saída:

theme	created_at	chat_id	description	owner_id
Discussion	2022-10-27 00:15:59.370000+0000	79dca637-b603-4b8c-81a1-bba4faf63b81	BLACK VS GREEN	d8c138ac-b35f-48db-8d6b-9e363cea2da2
Discussion	2022-10-27 00:19:39.060000+0000	f5c40ad9-d99f-43e6-a99c-96775cc945bc	ice cream or açaí	fdd2805c-1a70-48d9-8140-ff4066cf6c5e

WHERE e ORDER BY

```
SELECT * FROM example.chats_by_theme WHERE theme = 'Discussion'
ORDER BY created_at DESC;
```

Saída:

theme	created_at	chat_id	description	owner_id
Discussion	2022-10-27 00:19:39.060000+0000	f5c40ad9-d99f-43e6-a99c-96775cc945bc	ice cream or açaí	fdd2805c-1a70-48d9-8140-ff4066cf6c5e
Discussion	2022-10-27 00:15:59.370000+0000	79dca637-b603-4b8c-81a1-bba4faf63b81	BLACK VS GREEN	d8c138ac-b35f-48db-8d6b-9e363cea2da2

ALLOW FILTERING

```
SELECT * FROM example.chats_by_theme WHERE owner_from = 'BR' ALLOW
FILTERING;
```

Saída:

theme	created_at	chat_id	description	owner_from	owner_id
Blurt Out	2022-10-27 01:07:28.107000+0000	e12c042b-d5f8-4410-8374-9a9261b2ae23	rosas são vermelhas violetas são azuis	BR	ef458d9f-c0e0-4270-b5a6-b1057380d4a4

5) Crie pelo menos 1 índice secundário

```
-- Secondary Index

CREATE INDEX ON example.topic (tags);

SELECT * FROM example.topic WHERE tags CONTAINS 'news';
```

Saída:

topic_id	shares	tags	topic_name
459f789e-f264-4d75-958b-91c6d71ff6a2	{'facebook': 70, 'instagram': 100}	{'news', 'tragedy'}	dead man found in mansion
6b334761-4778-4e4e-8662-bdcd3330e938	{'facebook': 5, 'instagram': 20}	{'news'}	the queen lost her son and father

Referências Bibliográficas

[1] Creating collections. Disponível em:

https://docs.datastax.com/en/cql-oss/3.3/cql/cql_using/useCollections.html. Acesso em: 27 de outubro de 2022.

[2] The Cassandra Query Language (CQL). Disponível em:

<https://cassandra.apache.org/doc/latest/cassandra/cql/>. Acesso em: 27 de outubro de 2022.

[3] Get Started with Apache Cassandra. Disponível em:
https://cassandra.apache.org/_/quickstart.html. Acesso em: 27 de outubro de 2022.

[4] Intro to Apache Cassandra for Developers. Disponível em:
<https://github.com/datastaxdevs/workshop-intro-to-cassandra>. Acesso em: 27 de outubro de 2022.

[5] Modelo de Dados Colunar. Disponível em:
https://sig-arq.ufpb.br/arquivos/2022164169e0f1466458953273304549e/BigData_Lab04.pdf. Acesso em: 27 de outubro de 2022.

[6] Hibernated. Disponível em:
<https://docs.datastax.com/en/astra-serverless/docs/manage/db/manage-db-status.html>. Acesso em: 27 de outubro de 2022.