Astra DB

Big Data Systems

Dr. Rubi Boim

Motivation (for this course)

- Managing a cluster of Cassandra is not trivial
  Not the focus of this course

- Astra DB is a managed Cassandra DB service
  - Free version
  - We will use it to learn basic Cassandra programming

DataStax

- A commercial company established at 2010
  Cassandra was released at 2008

Provides

- Support for Cassandra
- DSE (DataStax Enterprise)
  own version of Cassandra with extra features
- Managed Database-as-a-service
  based on Apache Cassandra

DataStax Enterprise (DSE)

- Commercial product - NOT open source
  an extension of Cassandra

Provides extra features / tools on top of Cassandra

- Cluster management
- GraphQL support
- Analytics
- Search capabilities
- …

Astra DB

- A managed Cassandra database service
- Launched at 2020
- Support Cassandra +3.11

Astra DB can be configure to run on

- AWS
- GCP
- Azure

Create an account

https://astra.datastax.com/

Create an account

https://astra.datastax.com/

You do NOT need to enter credit card details (!!!)

DATASTAX

Start Today with $300/Year Credit

Sign Up with GitHub

Sign Up with Google

OR

First Name

Last Name

Company Email

I agree to DataStax's Privacy Notice, MSA and Terms of Service

Get Started Today

Already have an account? Sign In ›

› DATASTAX ASTRA DB

Build a Bold New World with Generative AI

One-stop GenAI Stack

Everything you need for RAG in one place — all the vector and structured data, tools, and integrations in an easy API that 'just works'.

Relevant GenAI FTW!

Up to 20% higher relevance, 9x more throughput, and 74x faster response time than standalone vector databases.*

Fast Path to Production

Built on the proven AI data leader. Planetary scale on any cloud with enterprise level security and compliance.

*Based on findings from Dec 2023 GigaOm Report: Vector Databases Compared v1.0

Uplevel your real-time AI journey alongside leaders in the industry

skypointcloud alpha ori TECHNOLOGIES bud priceline® dataworkz SupPlant

DataStax

Register

Sign Up With GitHub

Sign Up With Google

OR

First Name

Last Name

Work Email

Password

I agree to the DataStax MSA, including the Astra Supplement.

Create Account

Already have an account? Sign In

Astra DB

Multi-cloud DBaaS Built on Apache Cassandra™

- Start in minutes. No credit card required. Up to 80 GB free monthly.
- Build faster with REST, GraphQL, CQL and JSON/Document APIs.
- Deploy multi-tenant or dedicated databases on AWS, Azure, or GCP.

Introducing Astra Streaming Beta!

Astra Streaming is an open, multi-cloud event-streaming and data-streaming cloud service powered by Apache Pulsar™, now available within Astra when you sign up for an account.

Before the “GEN AI” :)

DataStax Astra

Live Chat | Rubi Boim

Welcome to Astra! Let's finish signing up.

Tell us more about you

To provide you with a better experience, we need to know a bit more about you. Afterwards, we may be able to customize your experience, tools, and content.

What's your primary goal in using Astra? *

Learning Cassandra

What is your primary programming language?

Java

What role best describes you?

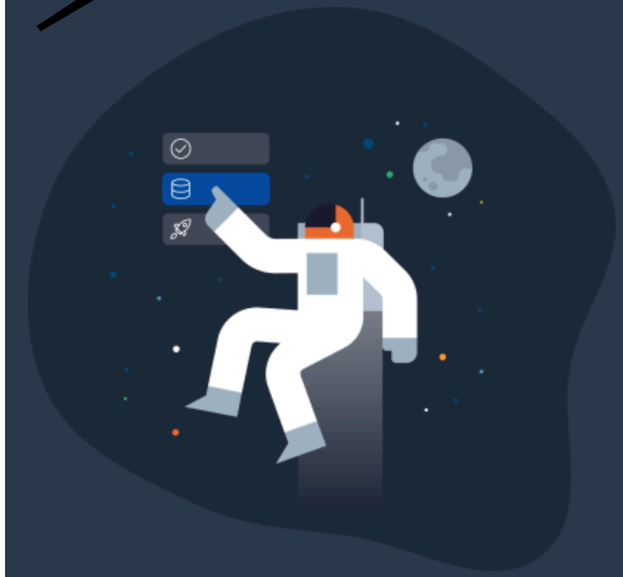
Student / Intern

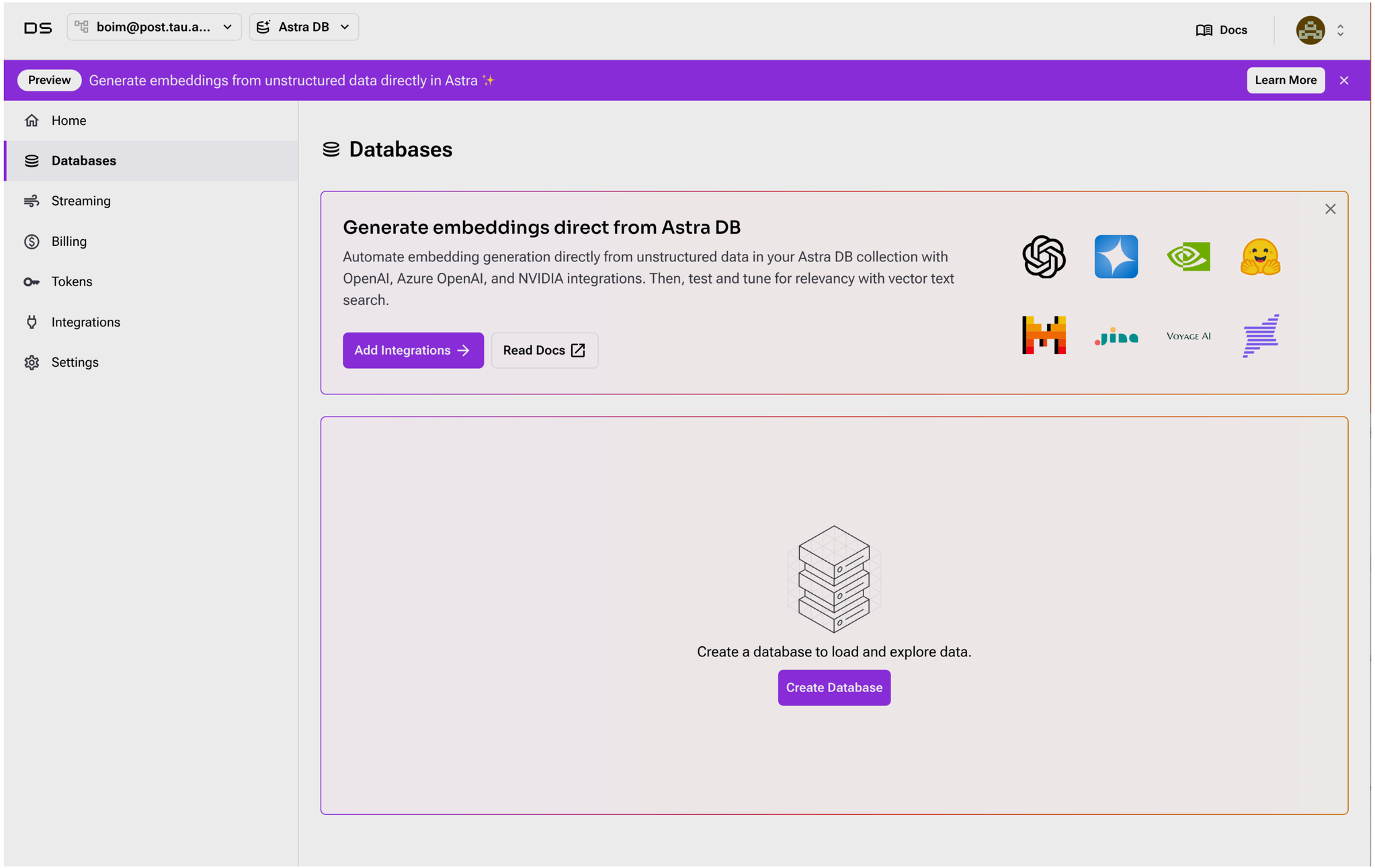
Are you currently running a Cassandra workload?

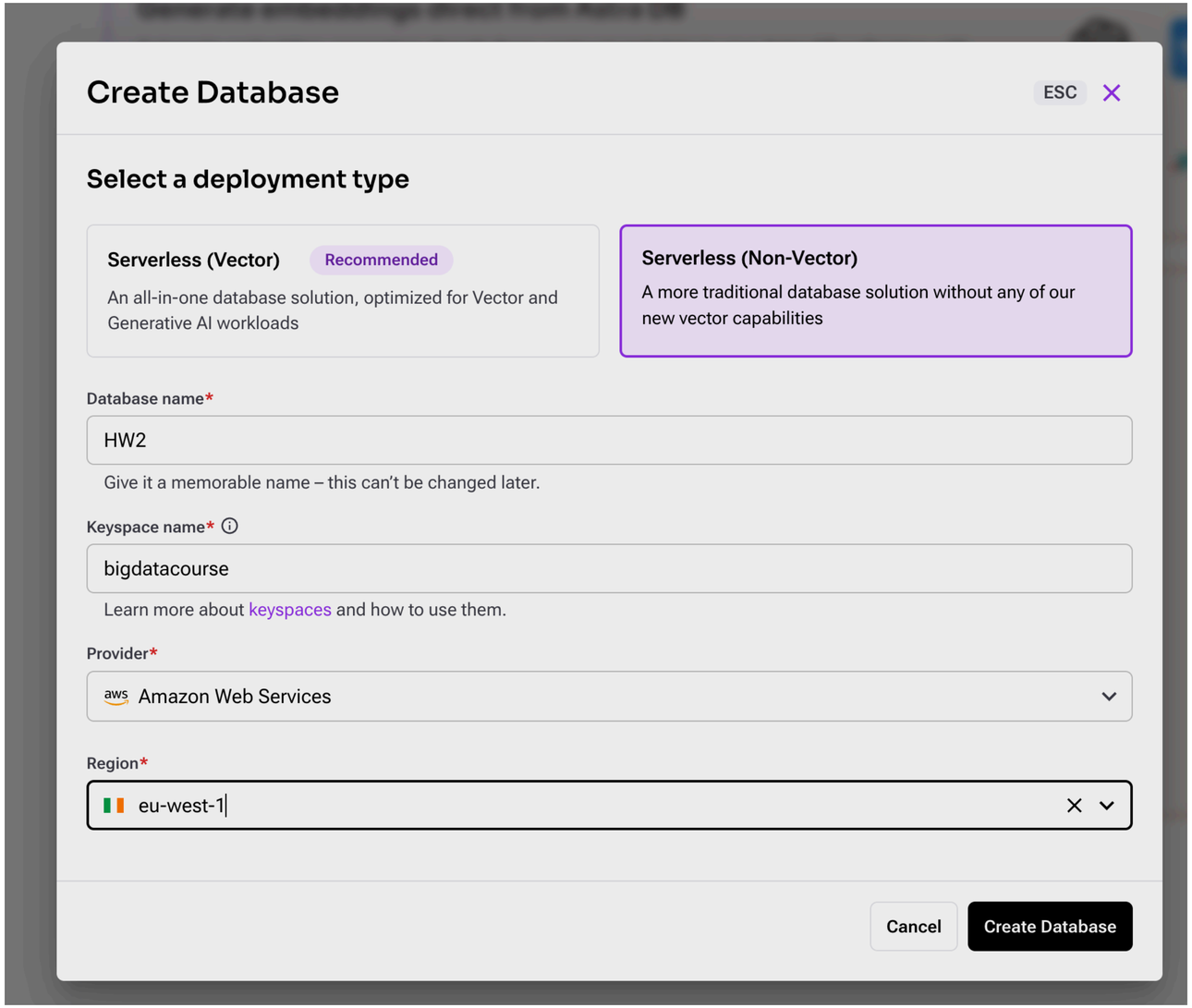
☐ Yes ☑ No

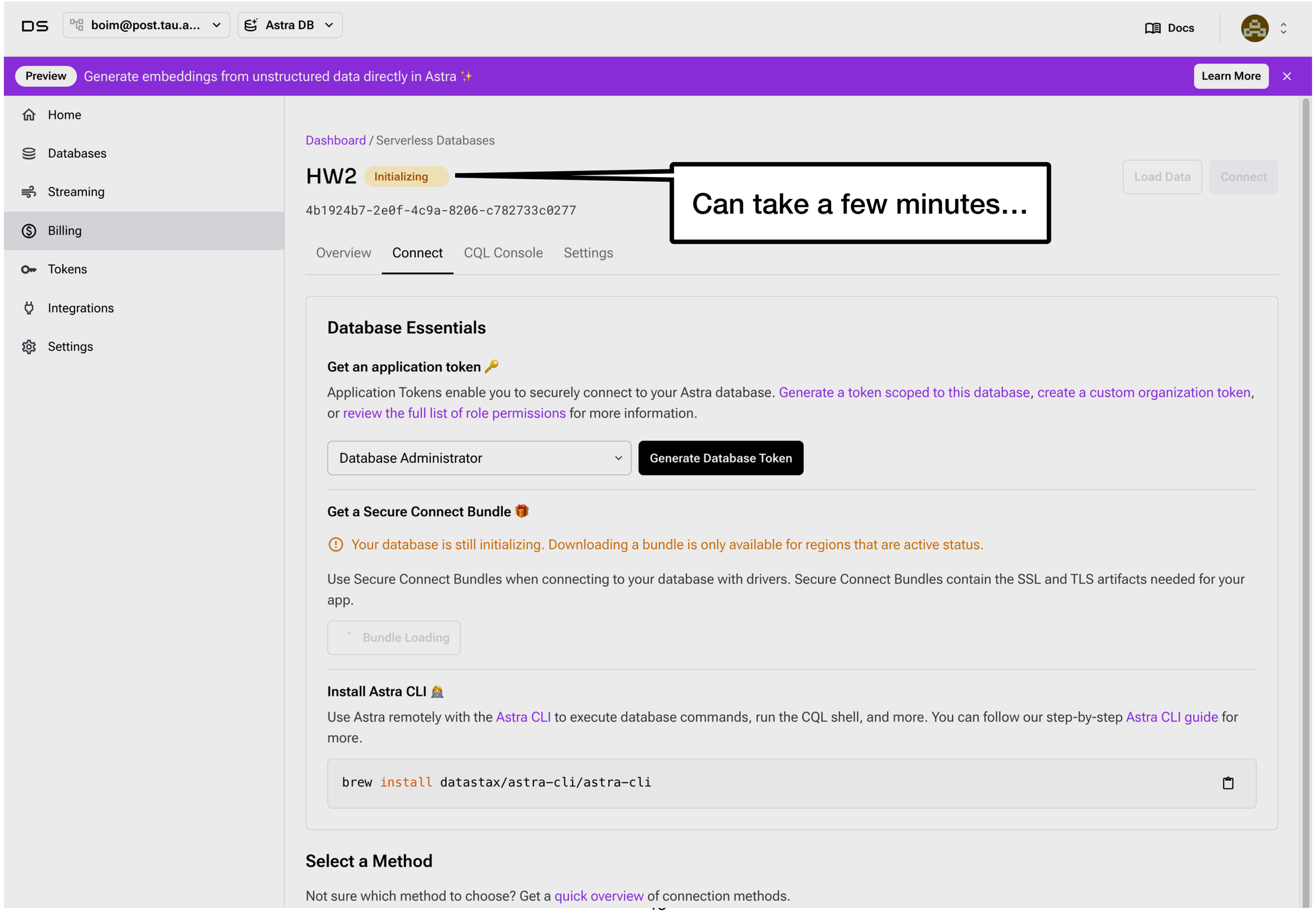
Next

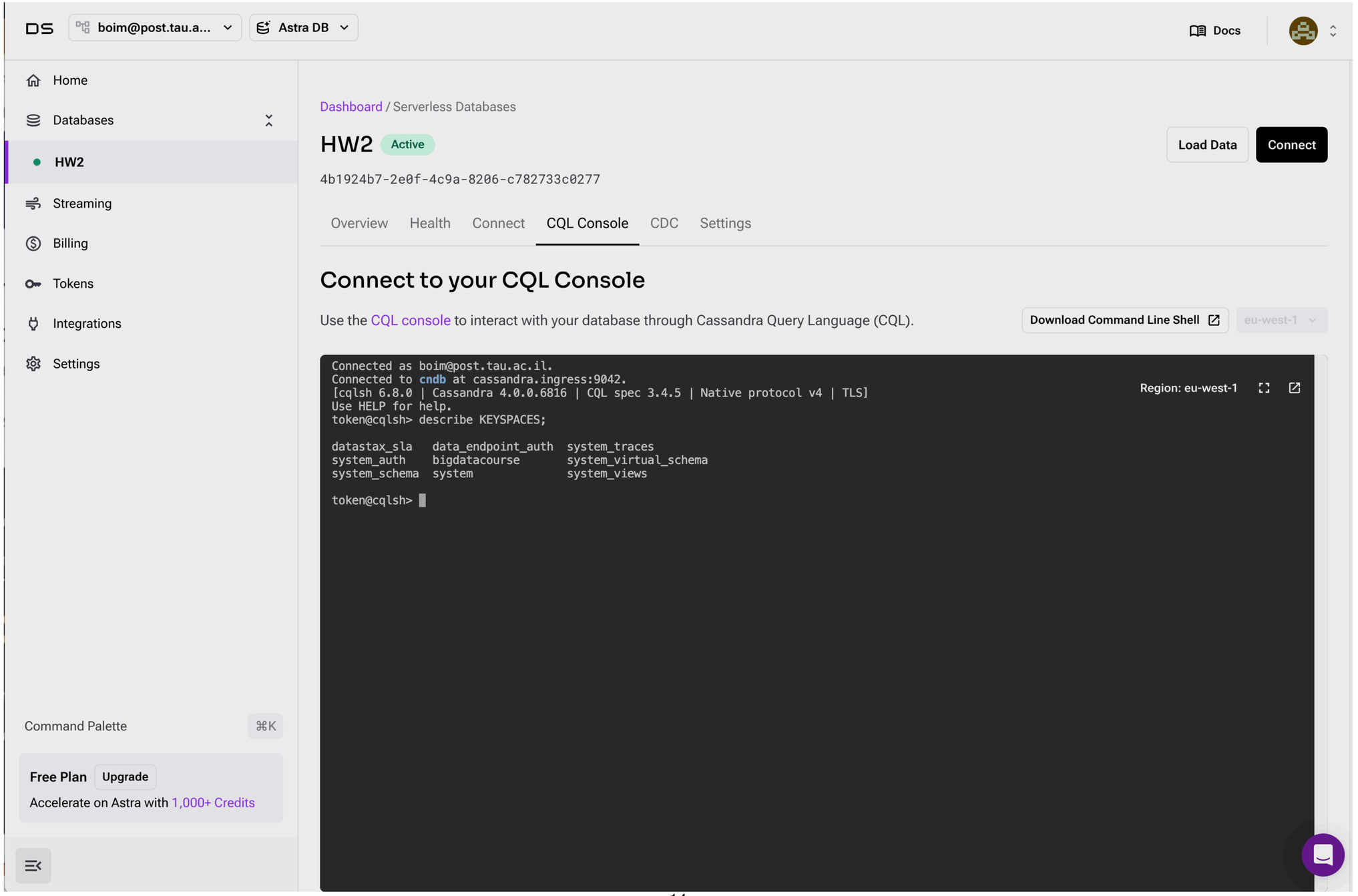
Not sure they still show this…

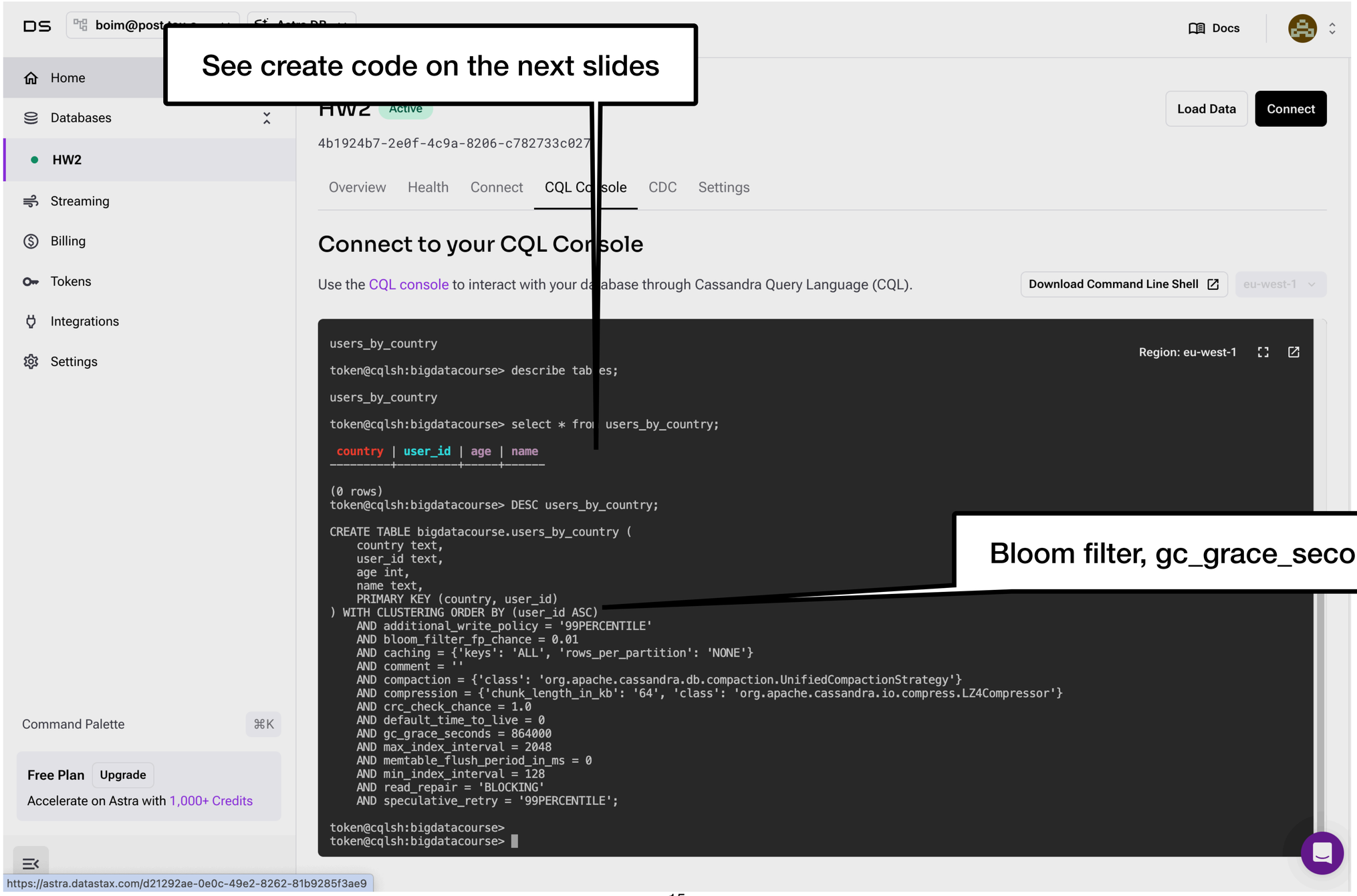



DS
boim@post.tau.a...
Astra DB
Docs
Preview
Generate embeddings from unstructured data directly in Astra ✨
Learn More
Home
Databases
Streaming
Billing
Tokens
Integrations
Settings
Databases
Generate embeddings direct from Astra DB
Automate embedding generation directly from unstructured data in your Astra DB collection with OpenAI, Azure OpenAI, and NVIDIA integrations. Then, test and tune for relevancy with vector text search.
Add Integrations
Read Docs
Voyage AI
Create a database to load and explore data.
Create Database


Create Database
ESC
Select a deployment type
Serverless (Vector)
Recommended
An all-in-one database solution, optimized for Vector and Generative AI workloads
Serverless (Non-Vector)
A more traditional database solution without any of our new vector capabilities
Database name*
HW2
Give it a memorable name – this can't be changed later.
Keyspace name*
bigdatacourse
Learn more about keyspaces and how to use them.
Provider*
Amazon Web Services
Region*
eu-west-1
Cancel
Create Database


boim@post.tau.a...
Astra DB
Docs
Preview
Generate embeddings from unstructured data directly in Astra ✨
Learn More
Home
Databases
Streaming
Billing
Tokens
Integrations
Settings
Dashboard / Serverless Databases
HW2
Initializing
4b1924b7-2e0f-4c9a-8206-c782733c0277
Can take a few minutes...
Load Data
Connect
Overview
Connect
CQL Console
Settings
Database Essentials
Get an application token 🔑
Application Tokens enable you to securely connect to your Astra database. Generate a token scoped to this database, create a custom organization token, or review the full list of role permissions for more information.
Database Administrator
Generate Database Token
Get a Secure Connect Bundle 🎁
Your database is still initializing. Downloading a bundle is only available for regions that are active status.
Use Secure Connect Bundles when connecting to your database with drivers. Secure Connect Bundles contain the SSL and TLS artifacts needed for your app.
Bundle Loading
Install Astra CLI 🦸
Use Astra remotely with the Astra CLI to execute database commands, run the CQL shell, and more. You can follow our step-by-step Astra CLI guide for more.
brew install datastax/astra-cli/astra-cli
Select a Method
Not sure which method to choose? Get a quick overview of connection methods.


DS
boim@post.tau.a...
Astra DB
Docs
Home
Databases
HW2
Streaming
Billing
Tokens
Integrations
Settings
Command Palette
⌘K
Free Plan
Upgrade
Accelerate on Astra with 1,000+ Credits
Dashboard / Serverless Databases
HW2
Active
Load Data
Connect
4b1924b7-2e0f-4c9a-8206-c782733c0277
Overview
Health
Connect
CQL Console
CDC
Settings
Connect to your CQL Console
Use the CQL console to interact with your database through Cassandra Query Language (CQL).
Download Command Line Shell
eu-west-1
Connected as boim@post.tau.ac.il.
Connected to cndb at cassandra.ingress:9042.
[cqlsh 6.8.0 | Cassandra 4.0.0.6816 | CQL spec 3.4.5 | Native protocol v4 | TLS]
Region: eu-west-1
Use HELP for help.
token@cqlsh> describe KEYSPACES;
datastax_sla data_endpoint_auth system_traces
system_auth bigdatacourse system_virtual_schema
system_schema system system_views
token@cqlsh>


See create code on the next slides
Home
Databases
HW2
Streaming
Billing
Tokens
Integrations
Settings
Command Palette
⌘K
Free Plan
Upgrade
Accelerate on Astra with 1,000+ Credits
Docs
HW2
Active
4b1924b7-2e0f-4c9a-8206-c782733c027
Load Data
Connect
Overview
Health
Connect
CQL Console
CDC
Settings
Connect to your CQL Console
Use the CQL console to interact with your database through Cassandra Query Language (CQL).
Download Command Line Shell
eu-west-1
Region: eu-west-1
users_by_country
token@cqlsh:bigdatacourse> describe tables;
users_by_country
token@cqlsh:bigdatacourse> select * from users_by_country;
country | user_id | age | name
(0 rows)
token@cqlsh:bigdatacourse> DESC users_by_country;
CREATE TABLE bigdatacourse.users_by_country (
country text,
user_id text,
age int,
name text,
PRIMARY KEY (country, user_id)
) WITH CLUSTERING ORDER BY (user_id ASC)
AND additional_write_policy = '99PERCENTILE'
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': '64', '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 = '99PERCENTILE';
token@cqlsh:bigdatacourse>
token@cqlsh:bigdatacourse>
Bloom filter, gc_grace_seconds, …
https://astra.datastax.com/d21292ae-0e0c-49e2-8262-81b9285f3ae9

Useful commands

- Autocomplete with the “tab” key
- `describe keyspaces;  (might work with “desc”)`
- `use <keyspace>`
- `describe tables`
- `describe <table>`
- `clear / CTRL+L (clear screen)`

- `create table / select / ...`

Demo - Create table

```
CREATE TABLE users_by_country (
   country              TEXT,
   user_id              TEXT,
   name                 TEXT,
   age                  INT,
   PRIMARY KEY ((country), user_id)
);
```

users_by_country	
country	K
user_id	▲C
name	
age	

Demo - Insert

users_by_country	
country	K
user_id	▲C
name	
age	

```
INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('Israel','user_123','Rubi Boim', 22);

INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('USA','user_123','Rubi Boim', 22);

INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('Israel','user_124','Tova Milo', 18);

INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('USA','user_125','Lebron James',35);

INSERT INTO users_by_country(country, user_id, name)
VALUES ('USA','user_125','Michael Jordan');
```

Demo - Insert

users_by_country	
country	K
user_id	▲C
name	
age	

```
INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('Israel','user_123','Rubi Boim', 22);

INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('USA','user_123','Rubi Boim', 22);

INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('Israel','user_124','Tova Milo', 18);

INSERT INTO users_by_country(country, user_id, name, age)
VALUES ('USA','user_125','Lebron James',35);

INSERT INTO users_by_country(country, user_id, name)
VALUES ('USA','user_125','Michael Jordan');
```

What is Michael's age?

Demo - Select

Which queries are valid?

users_by_country	
country	K
user_id	▲C
name	
age	

```
SELECT * FROM users_by_country;

SELECT * FROM users_by_country WHERE country = 'Israel';

SELECT * FROM users_by_country WHERE name = 'Rubi Boim';

SELECT * FROM users_by_country WHERE user_id = 'user_123';

SELECT * FROM users_by_country WHERE country = 'Israel' AND user_id =
'user_123';

SELECT * FROM users_by_country WHERE country = 'Israel' AND name =
'Rubi Boim';
```

Demo - Select

Which queries are valid?

users_by_country	
country	K
user_id	▲C
name	
age	

```
SELECT * FROM users_by_country;

SELECT * FROM users_by_country WHERE country = 'Israel';

SELECT * FROM users_by_country WHERE name = 'Rubi Boim';

SELECT * FROM users_by_country WHERE user_id = 'user_123';

SELECT * FROM users_by_country WHERE country = 'Israel' AND user_id =
'user_123';

SELECT * FROM users_by_country WHERE country = 'Israel' AND name =
'Rubi Boim';
```

Remember - “ALLOW FILTERING is ANTI PATTERN

Demo - Update

UPDATE is similar to INSERT
(*)

If you prefer to use the UPDATE syntax, “read the…”

Demo - Delete

Which queries are valid?

users_by_country	
country	K
user_id	▲C
name	
age	

```
DELETE FROM users_by_country WHERE name = 'Rubi Boim';

DELETE FROM users_by_country WHERE user_id = 'user_123';

DELETE FROM users_by_country WHERE country = 'Israel';

DELETE FROM users_by_country WHERE country = 'Israel'
AND name = 'Rubi Boim';

DELETE FROM users_by_country WHERE country = 'Israel'
AND user_id = 'user_123';
```

Demo - Drop table

```
DROP TABLE <table_name>;
```