

DataStax

Bring Streaming to Apache Cassandra with Apache Pulsar

Getting Started with Pulsar



LEVEL
UP
with the

DataStax
Developers

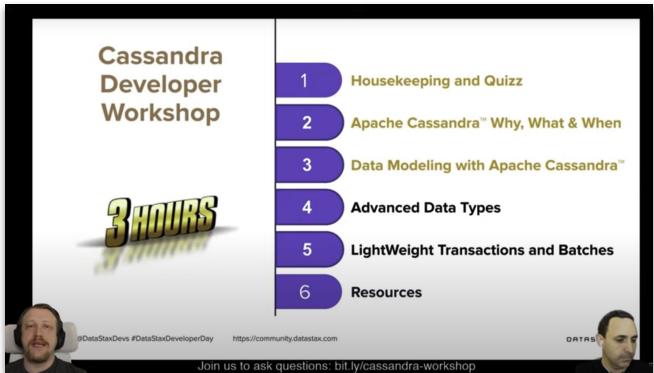
Guest Star : Jonathan Ellis

- Co-founder and CTO of **DataStax**
- Apache Cassandra Project Chair (PMC) for 7 years
- Expert in Distributed Systems
- Java Guru, also dangerous in Python
- Likes red shirts

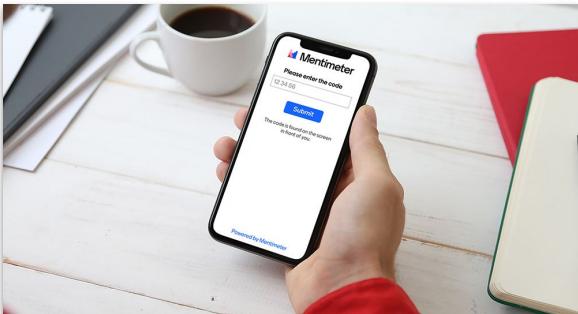
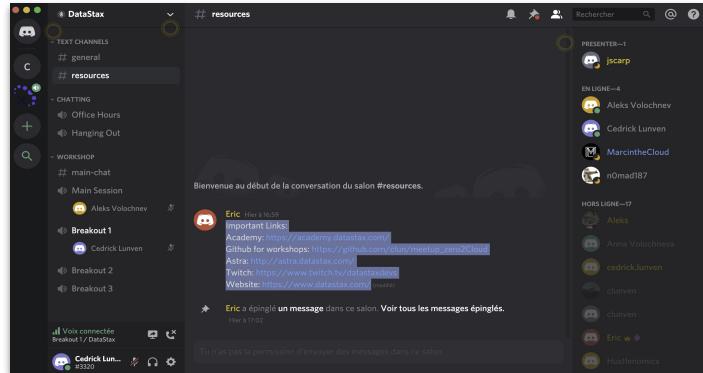


Housekeeping

Live: youtube.com/DataStaxDevs



Questions: bit.ly/cassandra-workshop



Quizz: menti.com ³



DataStax Developers

Labs

Astra DB interface showing a CQL console with the following text:

```
Connected to cass-cluster at stargate:9042...
[...] Use HELP for help.
token@qlsh:pulsar> describe keyspaces;
system_virtual_schemas system_auth system_traces system_distributed
system_data_stores system_views system_backups pulsar
system_schema system_data_endpoints stargate_system
token@qlsh:pulsar> use pulsar;
token@qlsh:pulsar> CREATE TABLE pulsar.products (
... id text PRIMARY KEY,
... description text
... );
token@qlsh:pulsar>
```

DataStax
Astra DB

KESQUE interface showing a dashboard with the following sections:

- Getting Started**: Describes Apache Kafka integration and provides links to GitHub samples.
- Cluster Locations**: Shows a world map with cluster locations and a "DATACENTER NAME" highlighted.
- Connecting**: Provides instructions for connecting using the Thrift binary protocol, Websocket API, or HTTP.



KESQUE

O'Reilly Katakoda Apache Pulsar™ Learning Series

By DataStax

Hands-on introduction to Apache Pulsar™

Getting Started with Apache Pulsar™

Creating Apache Pulsar™ Producers, Consumers and Readers in Java

Querying Data Streams with Apache Pulsar™

Connecting Apache Pulsar™ to Apache Cassandra™



File Edit Selection View Go Debug Terminal Workspace Help

EXPLORER: WORKSPACE

crud-nodejs

crud-nodejs

crud-python

db_connection.py

Ex02_Connect_to_Cass...

Ex03_Insert_Journey.py

Ex04_Travel.py

Ex05_Landing.py

Ex07_LatJourneys.py

Ex08_Read_Journey.py

Ex09_ReadMetrics.py

Ex10_ReadMetrics_Pag...

README.md

CRUD with NodeJS and ...

LICENSE

README.md

CRUD with Python and ...

LICENSE

gitpod /workspace/workspace/workshop-crud-with-python-and-node \$

gitpod /workspace/workspace/workshop-crud-with-python-and-node \$

Open Ports

3000 - not served

Wednesday, october 7th (NAM Time)

Thursday, october 8th (IST Time)

Items

gitpod /workspace/workspace/workshop-crud-with-python-and-node \$

gitpod /workspace/workspace/workshop-crud-with-python-and-node \$

Open Ports

3000 - not served

Make Private

CRUD operations with NodeJS and Python

Welcome to the CRUD operations with NodeJS and Python with Astra workshop! In this two-hour workshop, the Developer Advocate team of DataStax shows you the basics of connecting, updating and reading records from the powerful distributed NoSQL database Apache Cassandra.

Using Astra, the cloud based Cassandra-as-a-Service platform delivered by DataStax, we will cover the very first steps for every developer who wants to try to learn a new database: **CRUD operations**.

Video Recordings

- Wednesday, october 7th (NAM Time)
- Thursday, october 8th (IST Time)

DataStax Developers

menti.com

92 27 48 8



Available on the iPhone
App Store

GET IT ON
Google play

1 Introduction to Apache Pulsar

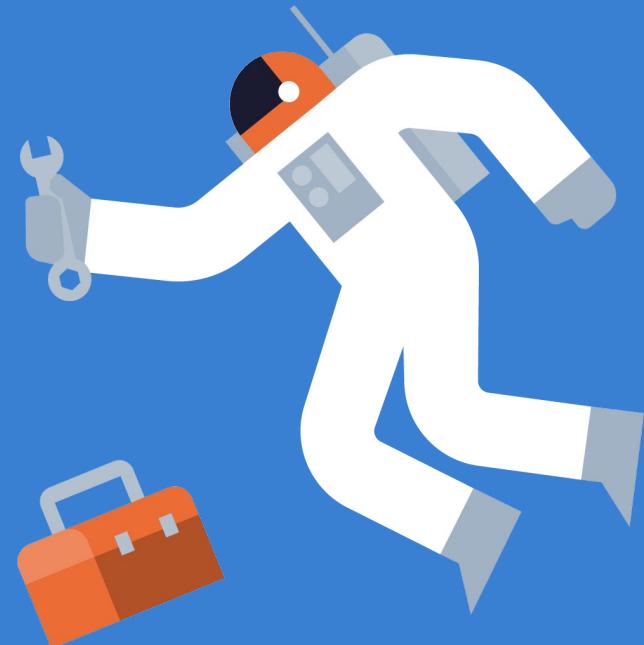
2 LAB #1 (katacoda)

3 LAB2 #2 (Kesque.com)

4 Apache Cassandra™ and Apache Pulsar™

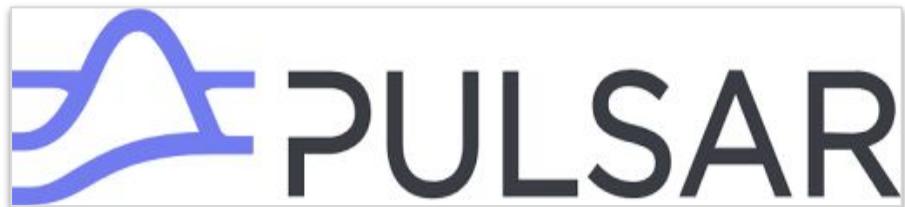
5 LABS 3 and 4 (Astra and Kesque)

6 Extra Resources and GAME !!



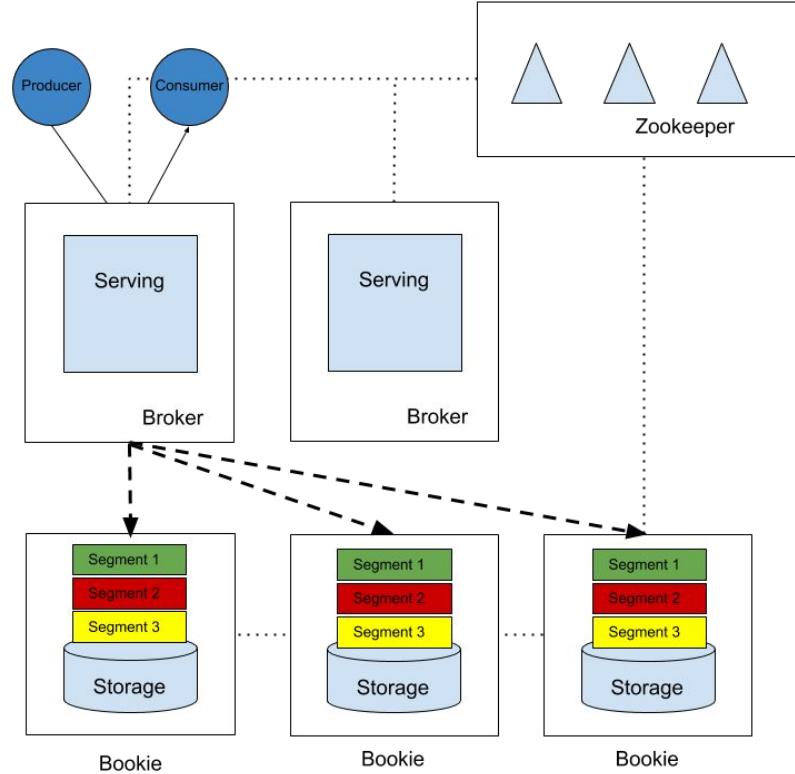
Apache Pulsar Overview

- First developed by **Yahoo**
- Open Source to Apache software foundation in **2016**
- Top-level Apache Project since **2018**
- **7k** stars on github, **350** contributors
- Streaming platform with pub/sub **and** queuing
- Split compute from storage
- Horizontal scalability / high performance / Low latency / Geo-replication



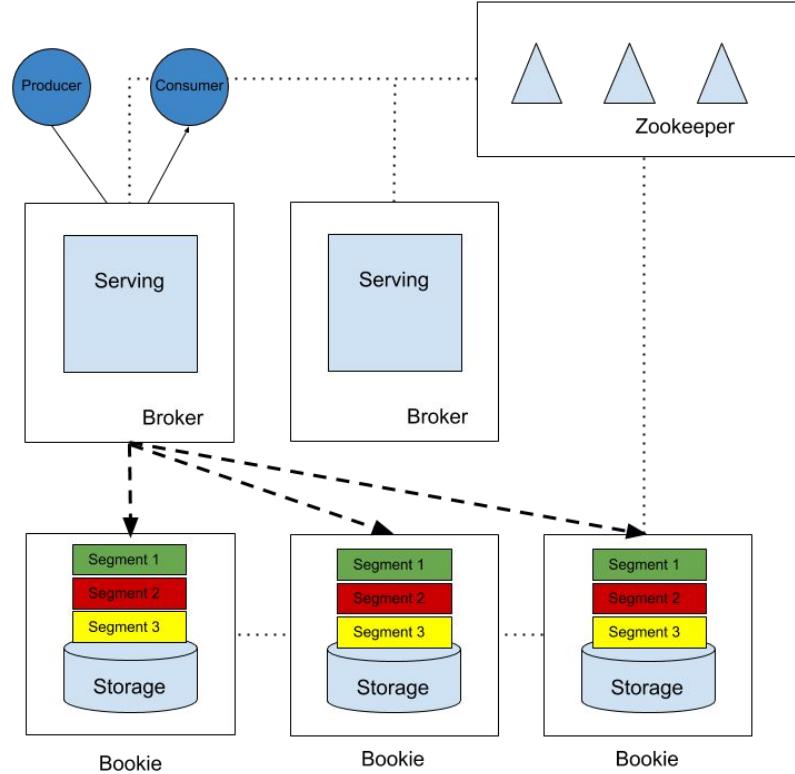
Architecture

- Distributed, tiered architecture
- Separates compute from storage
- Zookeeper holds metadata for the cluster
- Stateless Broker handles producers and consumers
- Storage is handled by Apache Bookkeeper



Architecture (cont'd)

- BookKeeper (“bookies”) distributed, append-only log
- Data is broken into ledgers and segments written to multiple bookies
- Producer acknowledged when quorum of bookies acknowledge
- No single bookie holds entire log

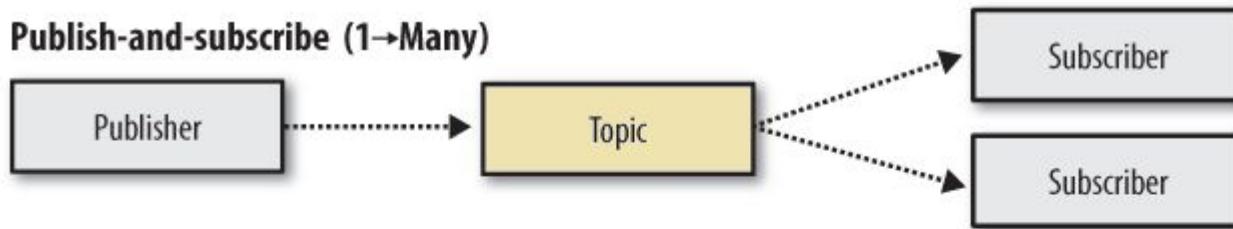


Why Pulsar?

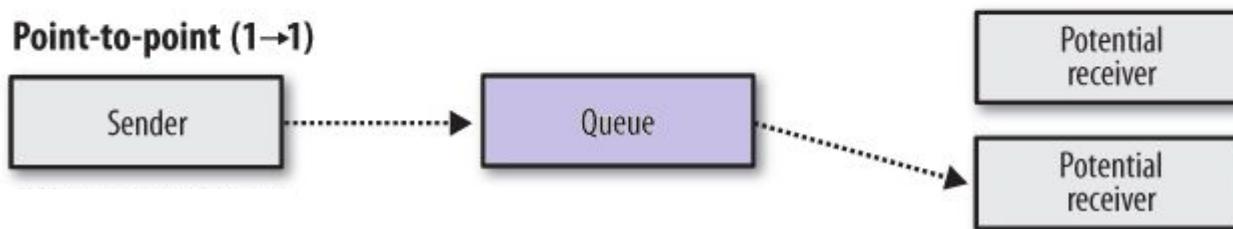
- Open source
- Pub/sub **and** queuing
- Geo-replication
- Multitenancy
- Better scaling

Pub/sub and queuing

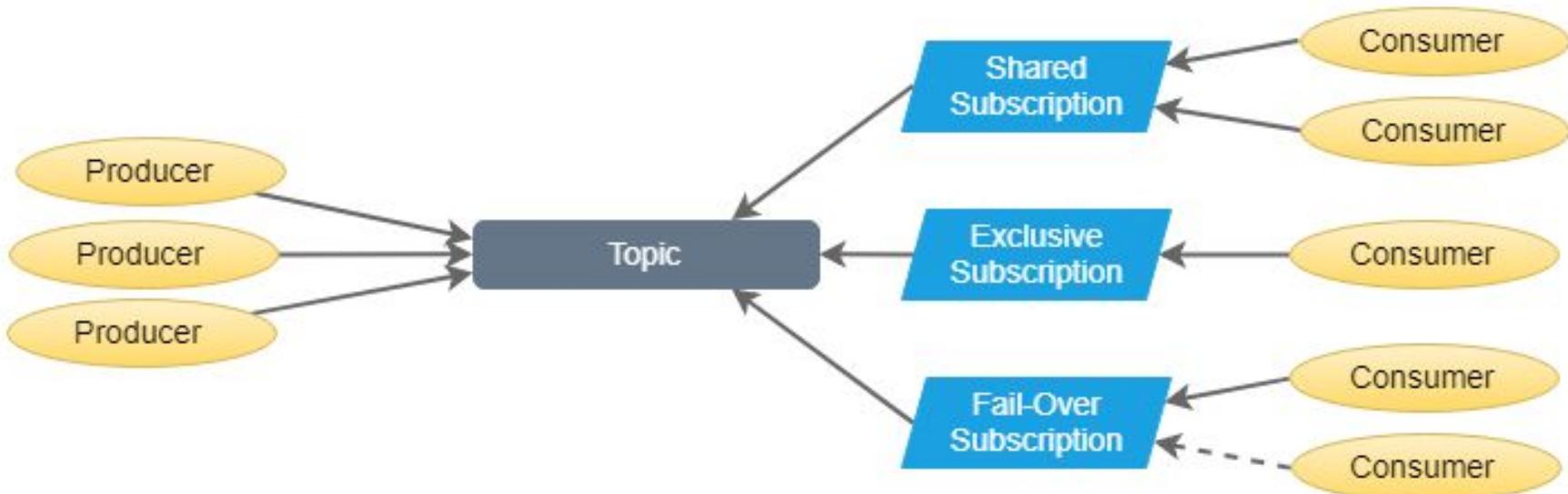
Publish-and-subscribe (1→Many)



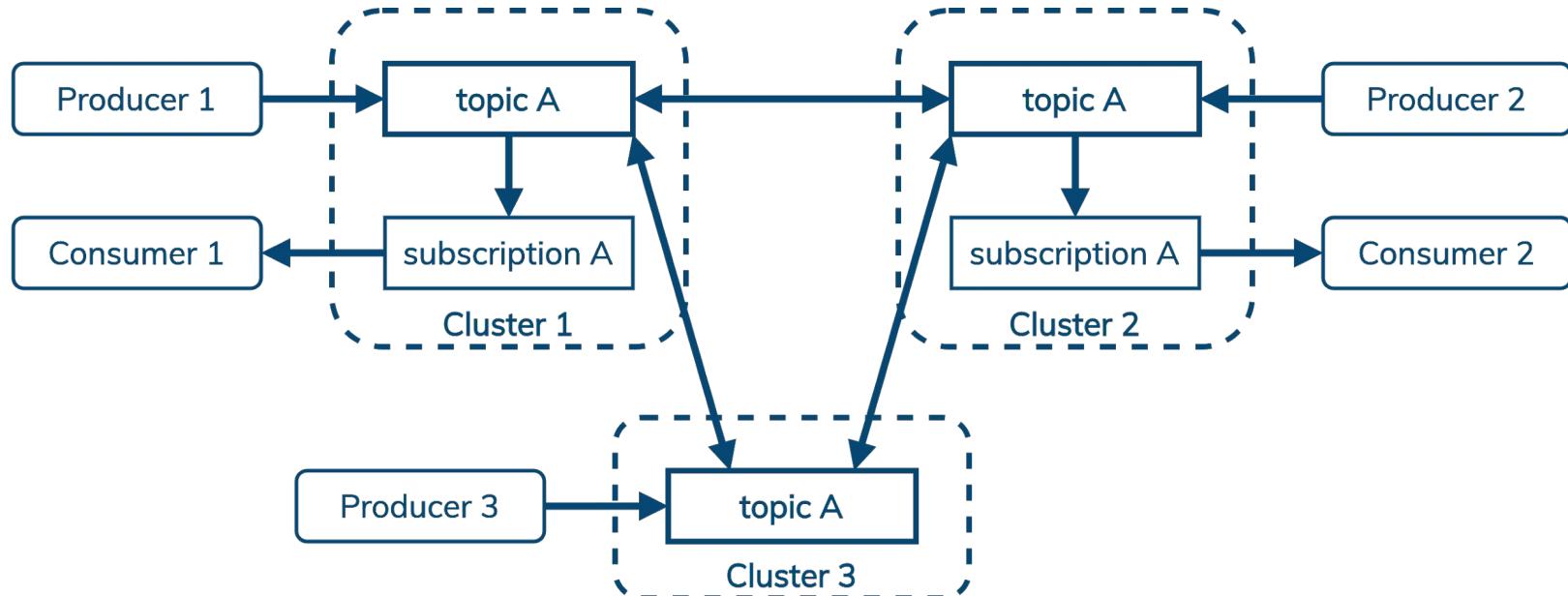
Point-to-point (1→1)



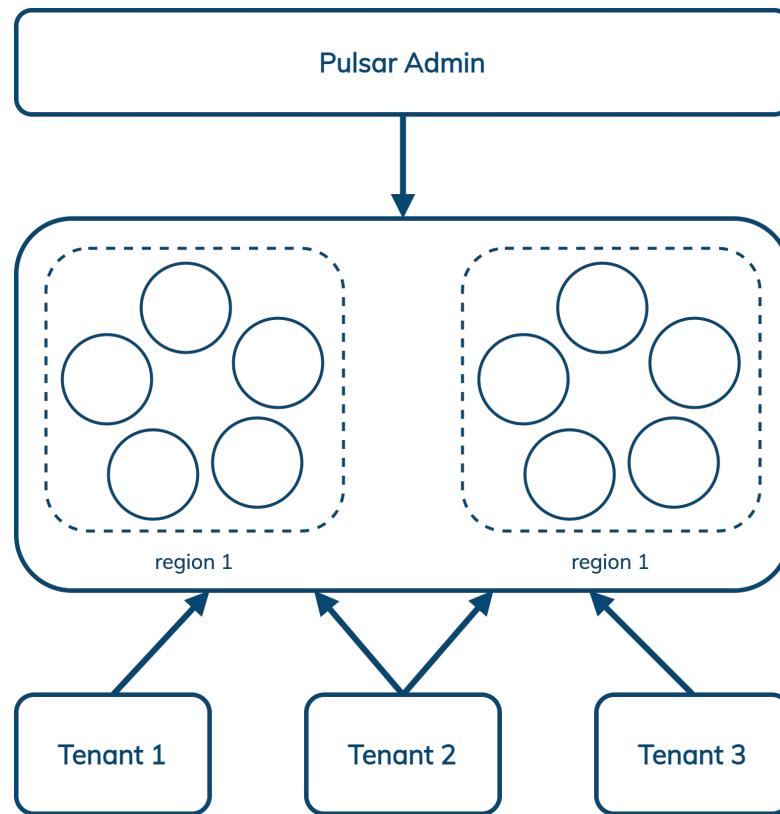
Pulsar implementation



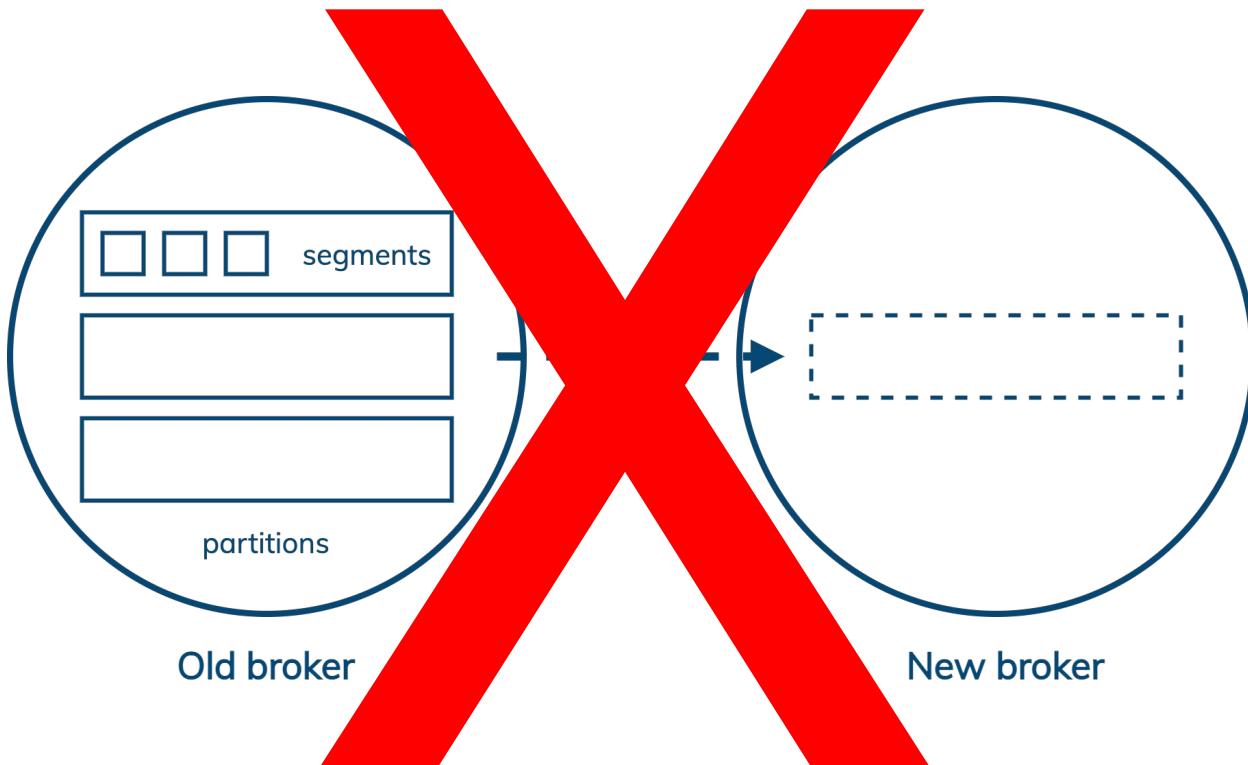
Geo Replication



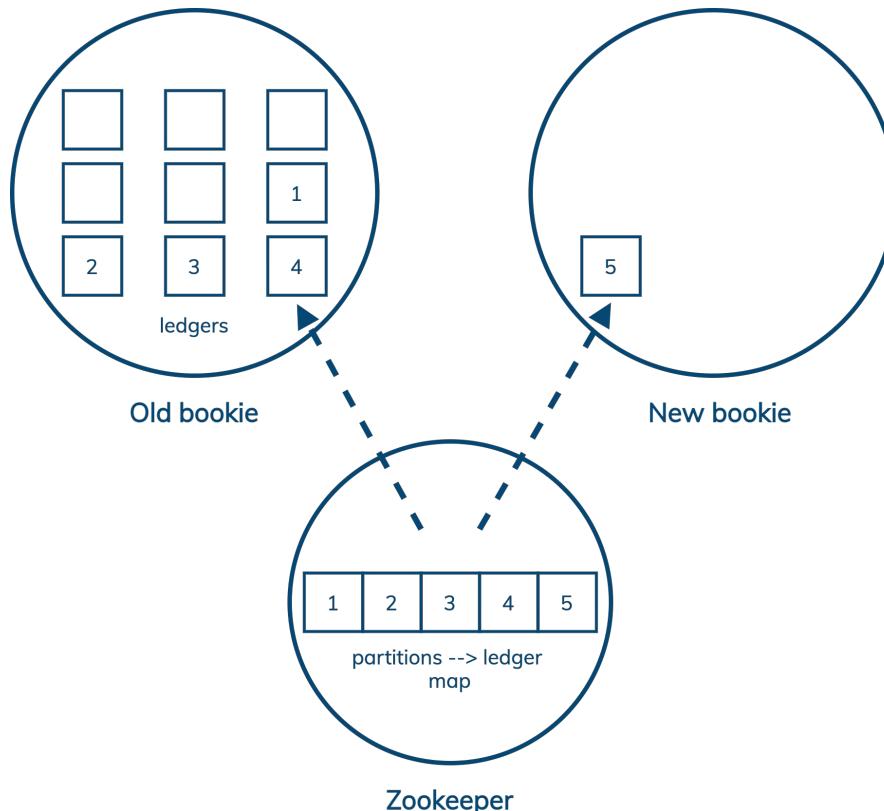
Multitenancy



Better scaling



Better scaling



1 Introduction to Apache Pulsar

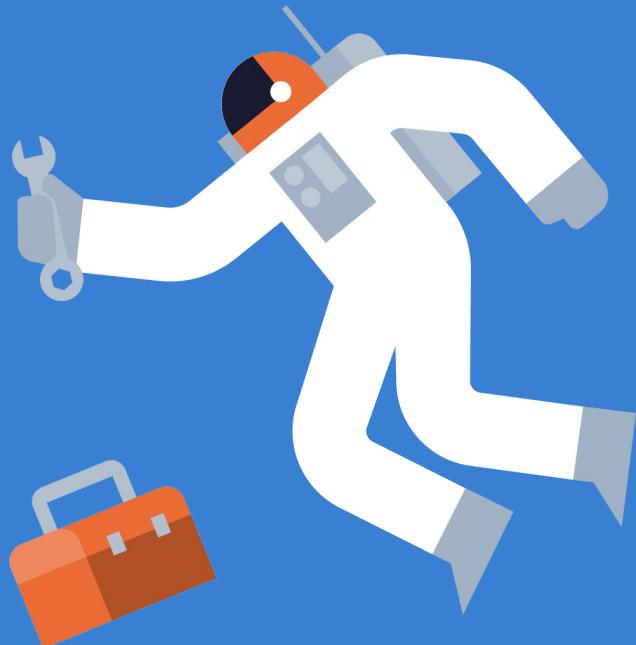
2 LAB #1 (katacoda)

3 LAB2 #2 (Kesque.com)

4 Apache Cassandra™ and Apache Pulsar™

5 LABS 3 and 4 (Astra and Kesque)

6 Extra Resources and GAME !!



LAB1 - 15 min



LAB1_GettingStarted
Cedrick Lunven edited this page 15 hours ago · 16 revisions

LAB1 - Getting started with Apache Pulsar™

 **PULSAR**

As described in the [official website](#), Apache Pulsar™ is a cloud-native, distributed messaging and streaming platform originally created at Yahoo! and now a top-level Apache Software Foundation project. To know more about the project you can browse the [reference documentation](#).

 **Katacoda**

During this first lab you will install Apache Pulsar from the tarball, create some infrastructure components, then a topic to store messages, then you will read and write messages from it. You don't have to install anything because we leverage Katacoda, a platform providing interactive learning and training for Software Engineers.

Before Starting

- 🚧 WARNING : Katacoda may not work on Brave browsers due to cookies being blocked. You can either use another browser like Chrome or Firefox or disable your ad-blockers.
- 👍 Let us know when you are done
Please give us a 👍 in the youtube chat when you are done.

Open Scenario

- Open the pulsar scenarios web page in a new tab of your browser (right-click or ctrl-click on Mac when opening the link) <https://www.katacoda.com/datastax/courses/pulsar>

▶ Pages 6

- Home**
- I - Getting Started
 - 1. Open Katacoda
 - 2. Start Scenario
 - 3. Run Scenario
- II - Kesque.com
 - 1. Create your account
 - 2. Create a Tenant
 - 3. Discover your cluster
 - 4. Create a topic
 - 5. Retrieve your token
 - 6. Create a producer
 - 7. Show messages in the UI
 - 8. Create a consumer
- III - Datastax Astra
 - 1. Create your account
 - 2. Create a Database
 - 3. Create a table
 - 4. Download the bundle
- IV - Event Streaming
 - 1. Create a sink
 - 2. Check sink in Kesque
 - 3. Check sink in Astra
- V - Resources

1 Introduction to Apache Pulsar

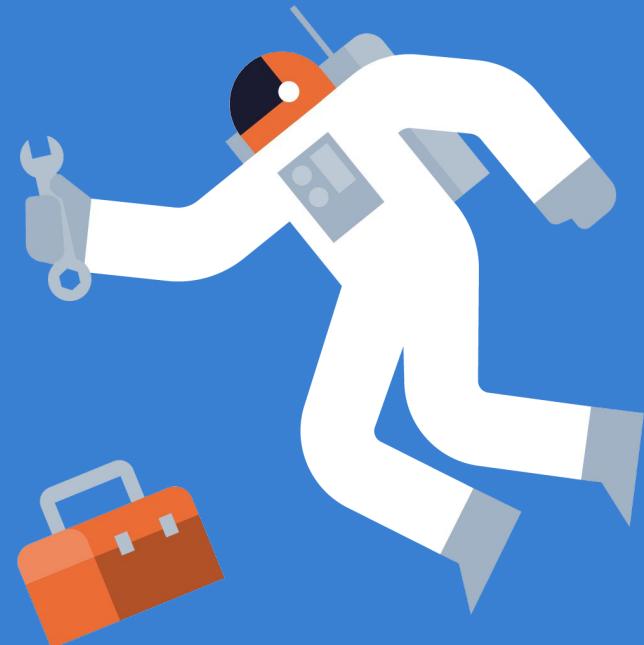
2 LAB #1 (katacoda)

3 LAB2 #2 (Kesque.com)

4 Apache Cassandra™ and Apache Pulsar™

5 LABS 3 and 4 (Astra and Kesque)

6 Extra Resources and GAME !!



DataStax Acquired Kesque



Subscription-to-Success with Open-Source Apache Pulsar™

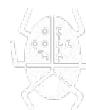
Pricing Plans

Frequently Asked Questions



SUPPORT

SINK



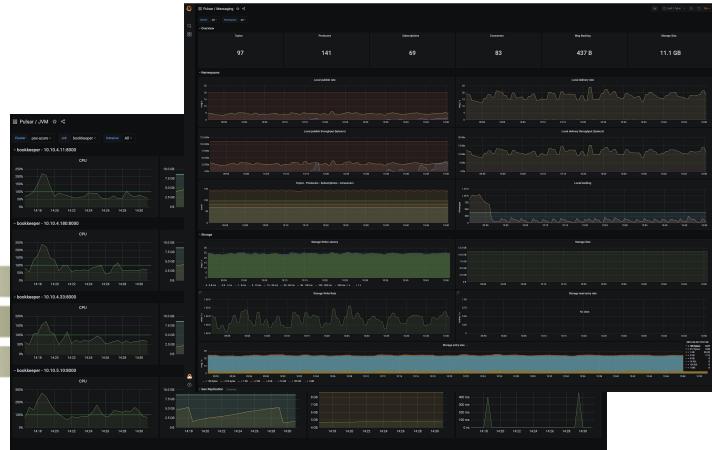
KESQUE

SAAS



kubernetes

Charts



Producer

- **Send messages to broker(s)**
- **A local message queue**
 - maxPendingMessages - queue size
 - queue management: working with broker using receipt
- **Message batching**
 - Batches tracked and stored as a single unit
 - Redelivery concern of unacked message

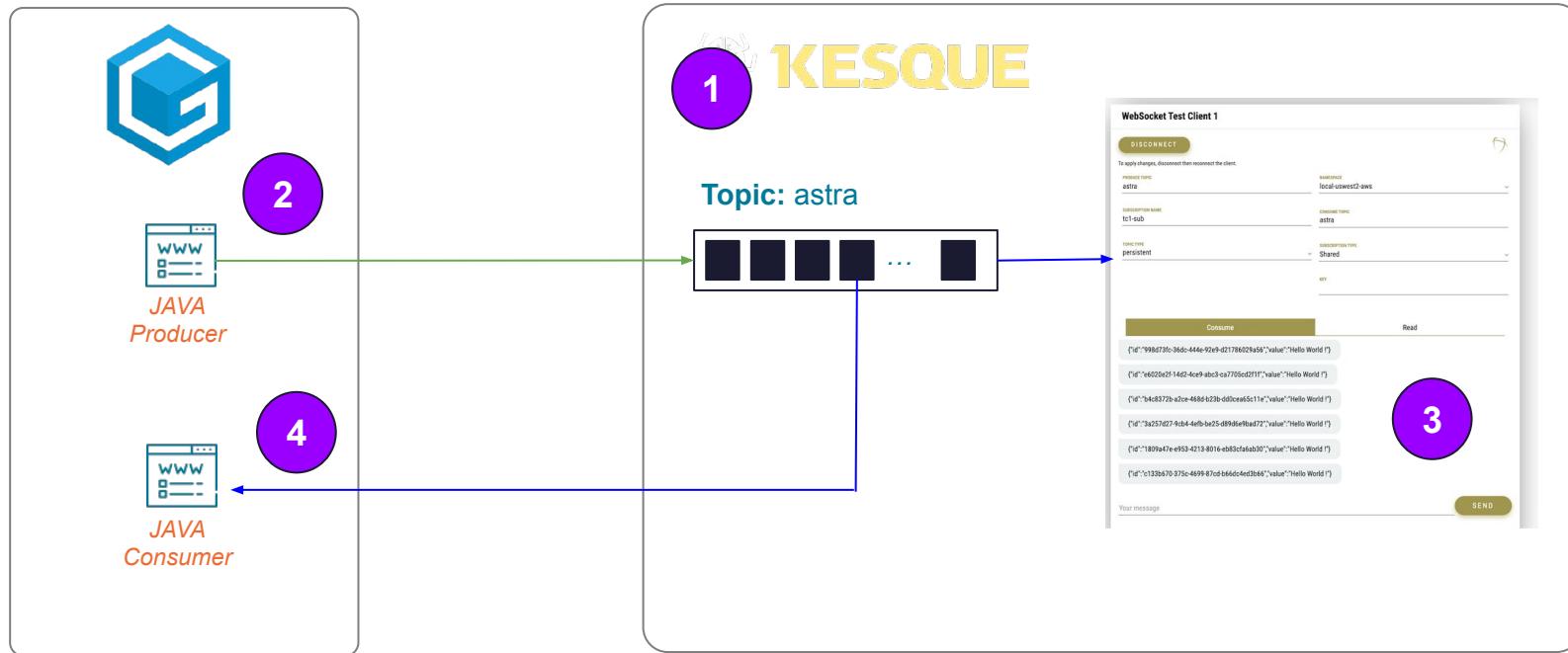
```
Producer<Data> producer =  
client.newProducer(JSONSchema.of(Data.class))  
    .topic("topic")  
    .create();  
producer.send(new Data(123));
```

Consumers and Subscriptions

- Send messages to broker(s)
- Consumer uses **subscription** tracks message consumption
- **Subscription cursor** is a reference pointer of message Id to track current consumption
 - use cursor to keep track of many things (replay, skip, rewind, seek, reader, subscription, ledger deletion, and etc.)

```
Consumer<Data> consumer = client  
  
.newConsumer(Schema.AVRO(Data.class)  
)  
.topic("topic")  
.create();  
consumer.receive();
```

LAB2 – What you will do



LAB2 - Kesque.com 40 min



KESQUE

LAB2_Kesque

Cedrick Lunven edited this page 13 hours ago · 22 revisions

LAB2 - Apache Pulsar™ as a Service with Kesque.com

KESQUE

STEP 1: Sign up to Kesque.com

<https://kesque.com> is a Cloud Messaging service, fully managed and Powered by Apache Pulsar®. It provides a free tier that you can use without providing any credit card.

Open <https://kesque.com> and locate the TRY FOR FREE button on the top right hand corner.

Kesque has been acquired by DataStax. Go [here](#) to find out more.



KESQUE is now KESQUE
Same superb service—newer name

HOME PRODUCT PRICING CONSULTING BLOG LOG IN TRY FREE

CLOUD MESSAGING SERVICE
Fully Managed, Powered by Apache Pulsar®

4 134 767 652
Average daily events we serve

Kesque is Available on Popular Cloud Services
Public, On-Prem, and Hybrid

aws Azure Google Cloud

Pages 6

Home

II - Getting Started

1. Open Katacoda
2. Start Scenario
3. Run Scenario

II - Kesque.com

1. Create your account
2. Create a Tenant
3. Discover your cluster
4. Create a topic
5. Retrieve your token
6. Create a producer
7. Show messages in the UI
8. Create a consumer

Datastax Astra

1. Create your account
2. Create a Database
3. Create a table
4. Download the bundle

IV - Event Streaming

1. Create a sink
2. Check sink in Kesque
3. Check sink in Astra

V - Resources

1. Slides
2. Keep in touch

A red circle highlights the 'Create your account' step under the 'Kesque.com' section of the sidebar.

1 Introduction to Apache Pulsar

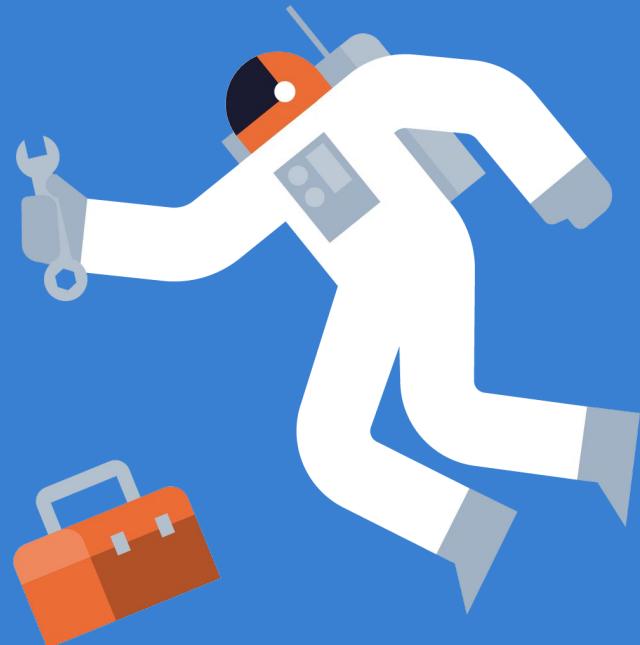
2 LAB #1 (katacoda)

3 LAB2 #2 (Kesque.com)

4 Apache Cassandra™ and Apache Pulsar™

5 LABS 3 and 4 (Astra and Kesque)

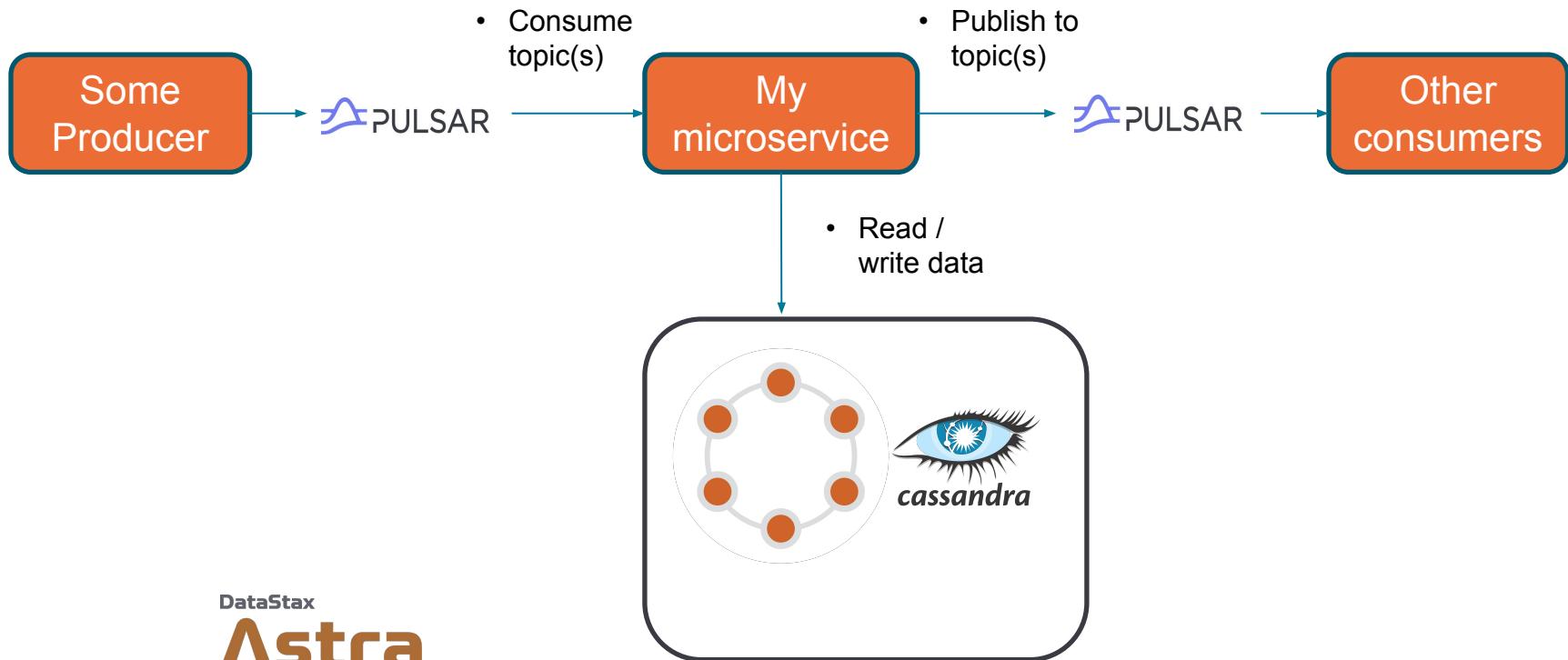
6 Extra Resources and GAME !!



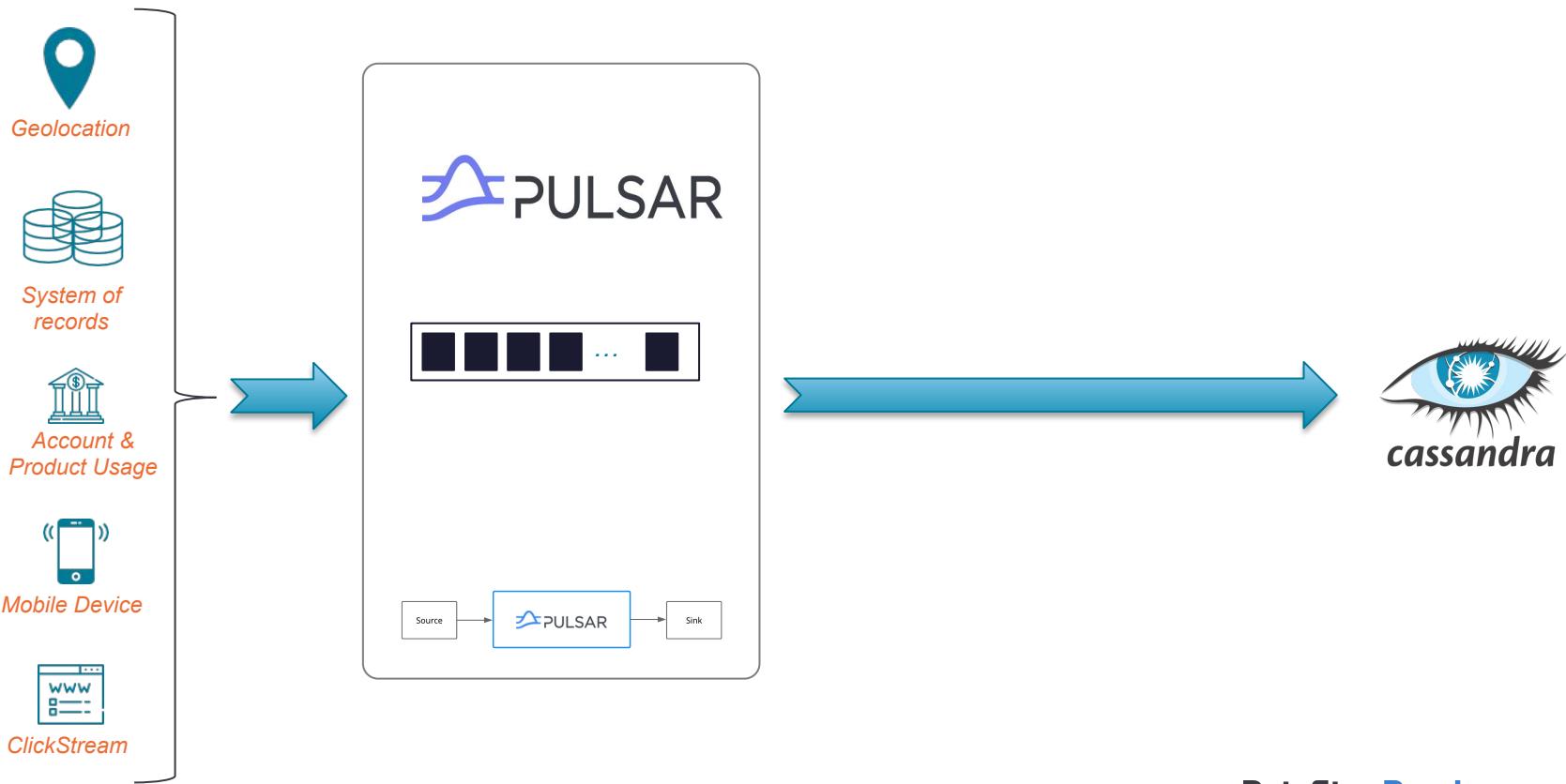
Cassandra + Pulsar – Similarities and Distinctives

- Concepts in common
 - Distributed Systems
 - Partitioning / Hashing
 - Replication
 - Slight differences in implementation
 - Multi-DC
 - Log-structured
 - TTL / retention
- Cassandra excels at...
 - High volume, write intensive data storage workloads at scale
 - Suitable as a system of record
 - High performance searching via DSE
- Pulsar excels at...
 - Streaming data to/from services and legacy data sources
 - Acting upon changes in data from multiple sources (aka pipelines)

Pattern 1: Cassandra + Pulsar in Microservices



Pattern 2: Event Streaming



Pulsar IO Connectors

- **Sources**
 - feed data from external systems into Pulsar.
- **Sinks**
 - feed data from Pulsar into external systems
- **When creating a connector, you can set the processing guarantee with the following semantics:**
 - ATLEAST_ONCE
 - ATMOST_ONCE
 - EFFECTIVELY_ONCE



Why a Pulsar Connector with Cassandra ?



Spark Streaming = PULL

- Computation framework
- Enable advanced transformations
- Mode Pull with a dedicated runtime (poll)



Pulsar IO Connector = PULL

- No extra runtime to consume
- Simple Mappings

1 Introduction to Apache Pulsar

2 LAB #1 (katacoda)

3 LAB2 #2 (Kesque.com)

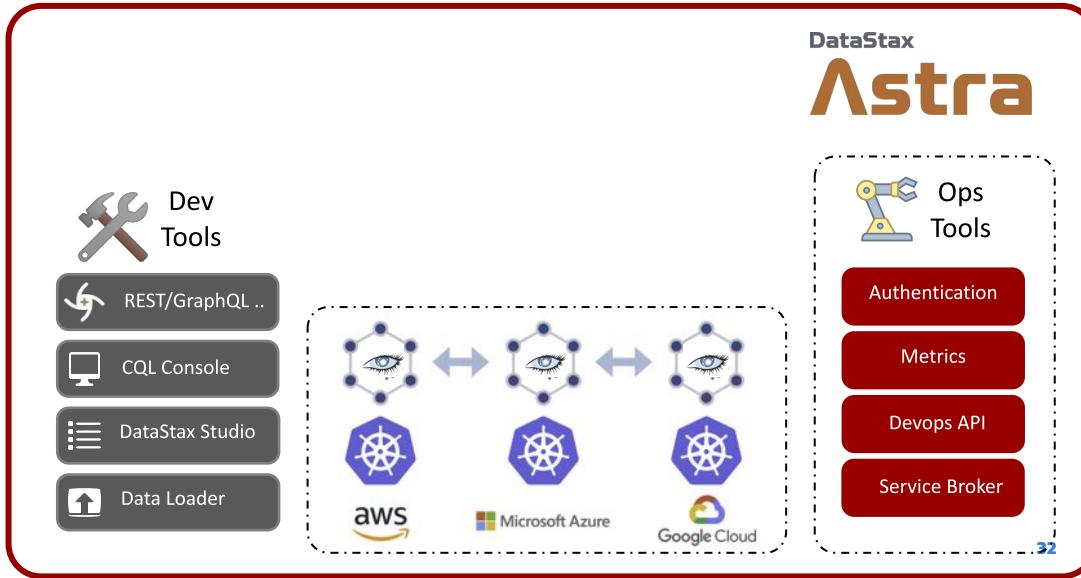
4 Apache Cassandra™ and Apache Pulsar™

5 LABS 3 and 4 (Astra and Kesque)

6 Extra Resources and GAME !!



Astra Platform



LAB3 - Astra 15 min



LAB3_Astra

David Jones-Gilardi edited this page 2 days ago - 8 revisions

Edit New Page

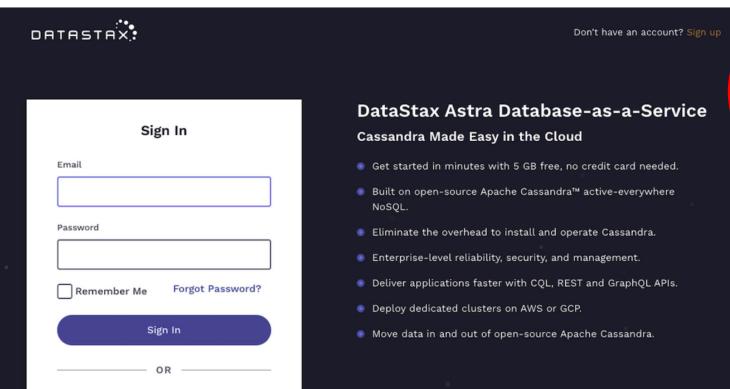
LAB3 - Saas Apache Cassandra™ with ASTRA

DataStax
Astra DB

STEP 1: Register to Astra Service

ASTRA service is available at url <https://astra.datastax.com>. ASTRA is the simplest way to run Cassandra with zero operations at all - just push the button and get your cluster. Astra offers 5 Gb Tier Free Forever and you don't need a credit card or anything to sign-up and use it.

• Register (if needed) and Sign In to Astra : You can use your Github, Google accounts or register with an email. Make sure to choose a password with minimum 8 characters, containing upper and lowercase letters, at least one number and special character



Don't have an account? [Sign up](#)

DataStax Astra Database-as-a-Service
Cassandra Made Easy in the Cloud

- Get started in minutes with 5 GB free, no credit card needed.
- Built on open-source Apache Cassandra™ active-everywhere NoSQL.
- Eliminate the overhead to install and operate Cassandra.
- Enterprise-level reliability, security, and management.
- Deliver applications faster with CQL, REST and GraphQL APIs.
- Deploy dedicated clusters on AWS or GCP.
- Move data in and out of open-source Apache Cassandra.

III - Datastax Astra

- Create your account
- Create a Database
- Create a table
- Download the bundle

IV - Event Streaming

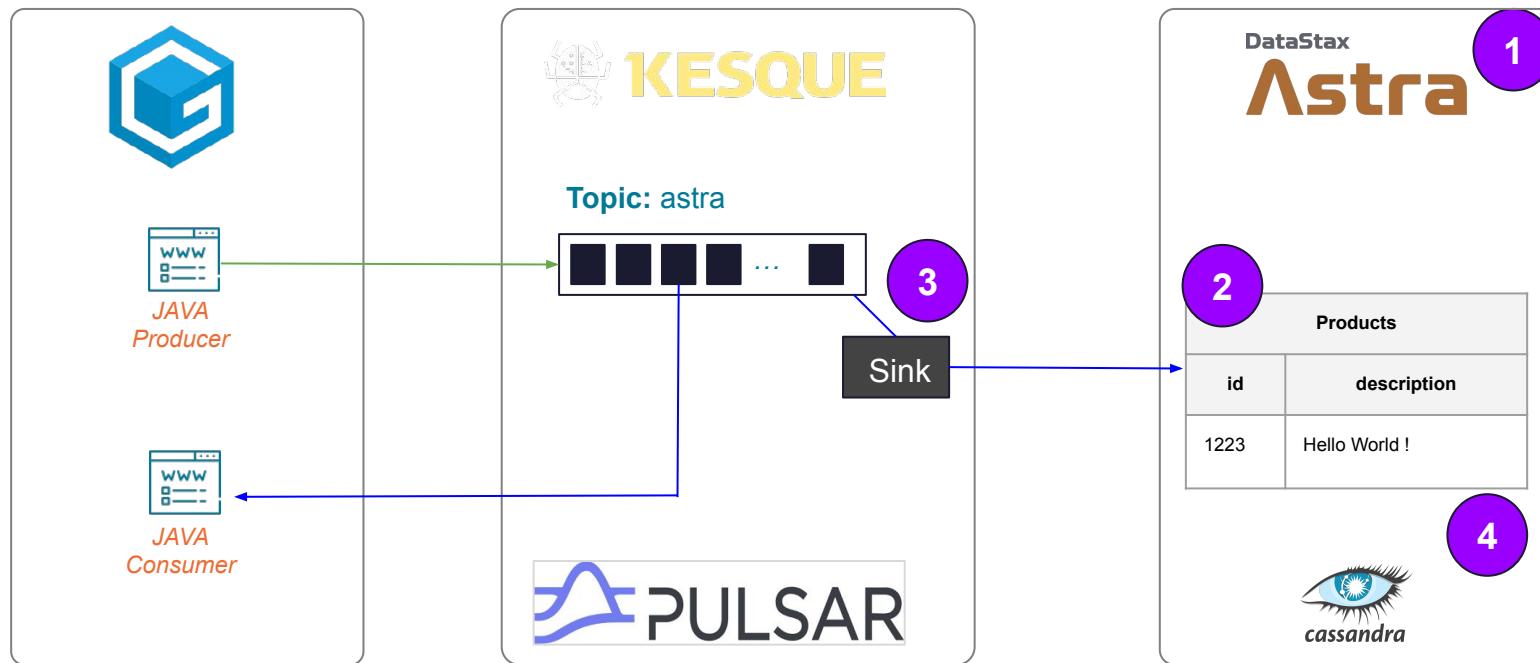
- Create a sink
- Check sink in Kesque
- Check sink in Astra

V - Resources

- Slides
- Keep in touch
- Keep Learning

Clone this wiki locally

LAB 3 & 4 – What you will do



LAB4 - Streaming

LAB4_Streaming

Cedrick Lunven edited this page 16 hours ago · 12 revisions

Edit

New Page

🎓 LAB4 - Streaming from Pulsar to Cassandra



✓ STEP 1: Create a sink

1. Select `Sinks > Add` in the menu on the left side
2. In the `SINKS` box select `cassandra-enhanced`, you might notice than Pulsar already manage a lot of different technologies. Today we will use the SINK implemented by Datastax.
3. In field `cloud.secureConnectBundle` upload the secureBundle zip file you downloaded on previous lab.
4. In field `auth.username` provide the astra user name `astraUser`
5. In field `auth.password` provide the astra user name `astraPassword`

This is what the user interface looks like for those first fields.

A screenshot of the KESQUE UI. On the left is a sidebar with 'Dashboard', 'Test Clients', 'Credentials', 'Code Samples', 'Topics', 'Functions', 'Sinks' (selected), and 'Manage'. Under 'Sinks', there are 'Add' and 'Manage' buttons. The main area shows a 'Basics' form with the following fields:

- SINKS:** A dropdown menu with 'cassandra-enhanced' highlighted.
- NAME:** 'cassandra-enhancedSink'
- FUNCTION:** 'local-userwest2.ows'
- SETTING:** 'PULSAR (REQUIRED)' dropdown set to 'contactPoints'.
- DESCRIPTION:** 'A DataStax Polar Sink to load records from Pulsar topics to Apache Cassandra(R) or DataStax Enterprise(SES)'.
- CLASS NAME:** 'com.datastax.oss.sink.pulsar.RecordCassandraSinkTask'

Annotations with arrows point to:

- #1 Select `Sinks > Add` (points to the 'Add' button in the sidebar).
- #2 Pick `Cassandra-enhanced sink` (points to the 'cassandra-enhanced' option in the dropdown).
- #3 Upload ZIP `secureBundle` (points to the 'contactPoints' setting in the form).

Pages 6

Home

I - Getting Started

1. Open Katacoda
2. Start Scenario
3. Run Scenario

II - Kesque.com

1. Create your account
2. Create a Tenant
3. Discover your cluster
4. Create a topic
5. Retrieve your token
6. Create a producer
7. Show messages in the UI
8. Create a consumer

III - DataStax Astra

1. Create your account
2. Create a Database
3. Create a table
4. Download the bundle

IV - Event Streaming

1. Create a sink
2. Check sink in Kesque
3. Check sink in Astra

V - Resources

1. Slides
2. Keep in touch
3. Keep Learning

1 Introduction to Apache Pulsar

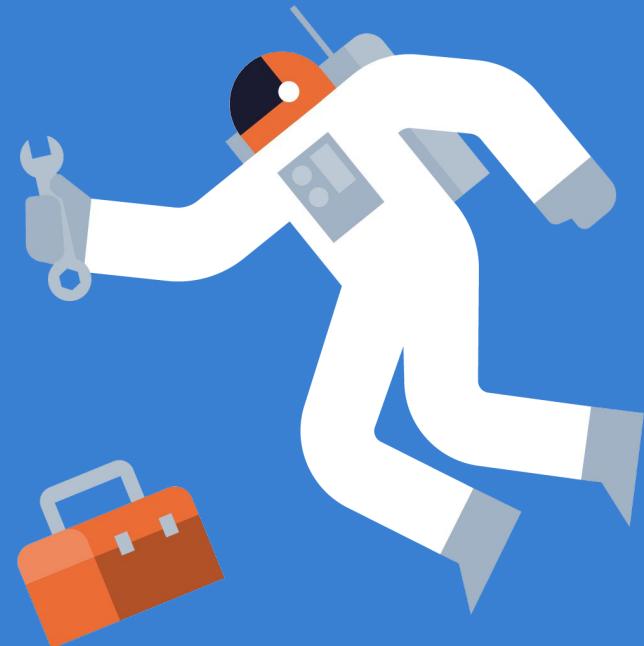
2 LAB #1 (katacoda)

3 LAB2 #2 (Kesque.com)

4 Apache Cassandra™ and Apache Pulsar™

5 LABS 3 and 4 (Astra and Kesque)

6 Extra Resources and GAME !!



Resources to Share

LAB5_Resources

Cedrick Lunven edited this page 14 hours ago - 10 revisions

Edit New Page

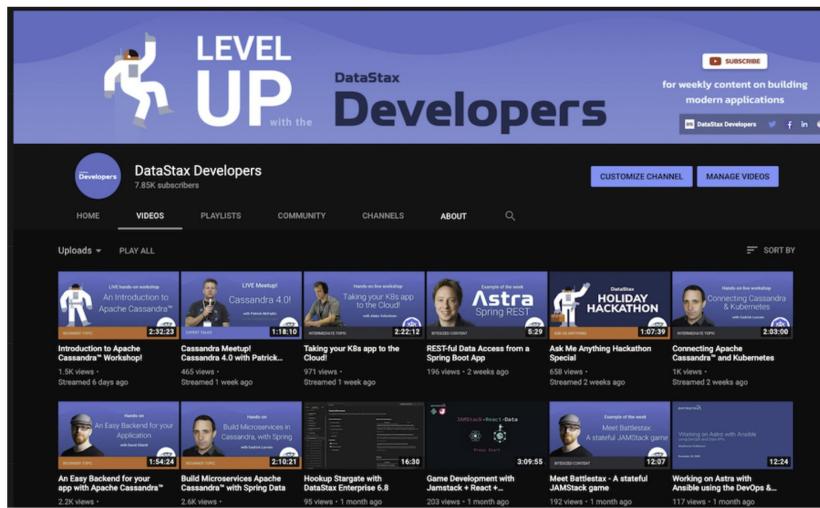
Extra Resources an links

Slides

TBD

👉 Let's keep in touch

🔗 DataStaxdevs Youtube Channel



▶ Pages 6

Home

» I - Getting Started

1. Open Katacoda
2. Start Scenario
3. Run Scenario

» II - Kesque.com

1. Create your account
2. Create a Tenant
3. Discover your cluster
4. Create a topic
5. Retrieve your token
6. Create a producer
7. Show messages in the UI
8. Create a consumer

» III - Datastax Astra

1. Create your account
2. Create a Database
3. Create a table
4. Download the bundle

» IV - Event Streaming

1. Create a sink
2. Check sink in Kesque
3. Check sink in Astra

» V - Resources

1. Slides
2. Keep in touch
3. Keep Learning

Pulsar Webinar 2/16th

[https://www.datastax.com/resources/webinar/
how-apache-pulsar-can-help-you-build-c-applications-nam](https://www.datastax.com/resources/webinar/how-apache-pulsar-can-help-you-build-c-applications-nam)





The New global Developer Economics survey is now open. Have your say on Dev Trends for 2021!

- Answer fun questions about the type of projects you're working on, your favourite languages, tools, technologies or platforms.
- In return, the survey will reward you with prize draws worth in total over \$17,000 USD. You can win new dev gear to upgrade your workstation, courses and licenses to learn something new, Amazon vouchers, and more.

Every developer who completes the survey will get a free virtual goodie bag!



<http://bit.ly/DEVSURVEY>
Find the link in the description below

Live Coding on YouTube

- Live coding sessions with advocates and guests
- Working through the challenges of building distributed systems
- Join the conversation and ask questions

The banner features a blue background with a white astronaut icon on the left. To the right of the icon, the text "LEVEL UP with the" is written vertically. Below this, the "DataStax Developers" logo is displayed, with "DataStax" in smaller blue text above "Developers" in large, bold, dark blue text. On the far right, there is a "SUBSCRIBE" button with a YouTube icon, and below it, text encouraging viewers to "for weekly content on building modern applications".



DataStax Developers

7.08K subscribers

HOME

VIDEOS

PLAYLISTS

COMMUNITY

CHANNELS

ABOUT





LEARN

New hands-on learning at www.datastax.com/dev
Classic courses available at DataStax Academy

ASK/SHARE

Join community.datastax.com
Ask/answer community user questions – share your expertise

CONNECT

Follow us @DataStaxDevs
We are on Youtube – Twitter – Twitch!

MATERIALS

Slides and exercises for this workshop are ⁴ available at
<https://github.com/DataStax-Academy/workshop-crud-with-python-and-node>

menti.com

92 27 48 8

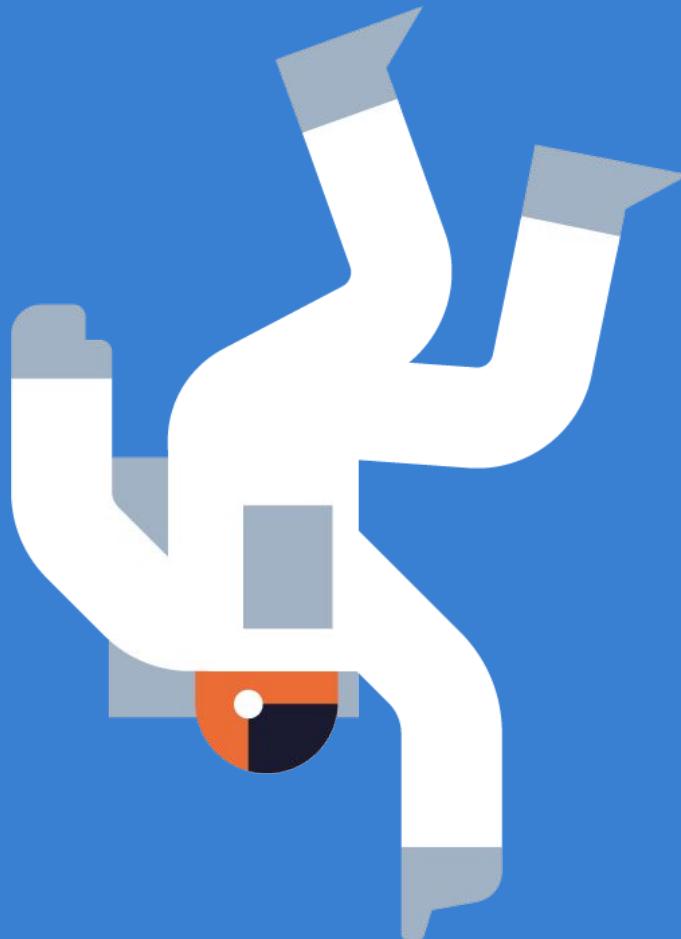


Available on the iPhone
App Store

GET IT ON
Google play

Giveaways!

As a thank you for joining us live, we would like to offer you the choice of giveaways



Option 1: Certifications

<https://www.datastax.com/dev/certifications>



A dark blue rectangular box containing promotional text. At the top right, it says "COMING SOON! Apache Cassandra Operations in Kubernetes Certification". Below that, a paragraph explains the program's purpose: "As teams work to containerize and deploy applications using Kubernetes, there's increasing interest in running Cassandra in Kubernetes alongside applications as well. We're developing a new certification program to help teams level up their skills to run Cassandra successfully in cloud-native deployments." It also mentions "The Apache Cassandra Operations in Kubernetes Certification will cover: running Cassandra in Docker containers, understanding how Cassandra maps to Kubernetes, and how to deploy and run Cassandra on Kubernetes using Kubernetes operators and other monitoring and management tools." At the bottom, it says "Sound interesting? Sign up to get notified about this program." and has a "SIGN UP NOW" button.

Vouchers (145\$ each) , valid 3 months, with 2 attempts included

Claim using the form provided in the chat

Option 2: \$300 Astra Code



Continue your Cassandra Dev journey! Your \$300 Astra credits will power an A5 database for five months or the more powerful A10 for one month.

Claim using the form provided in the chat

Thank you

