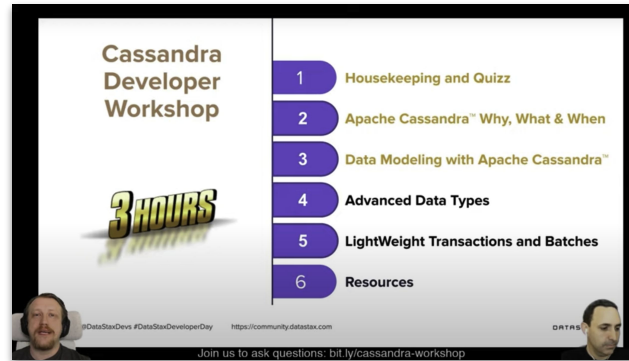




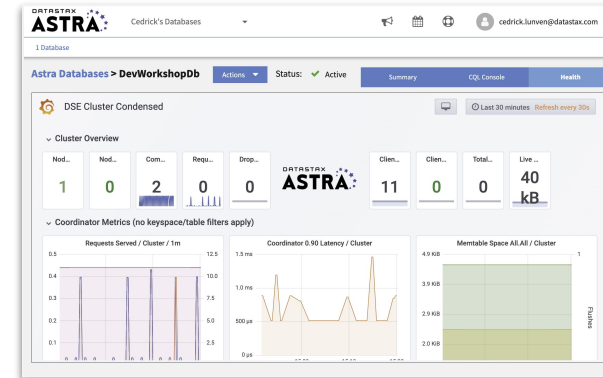
Welcome

**Supercharge your Cassandra
Data Model with new scalable
indexing**

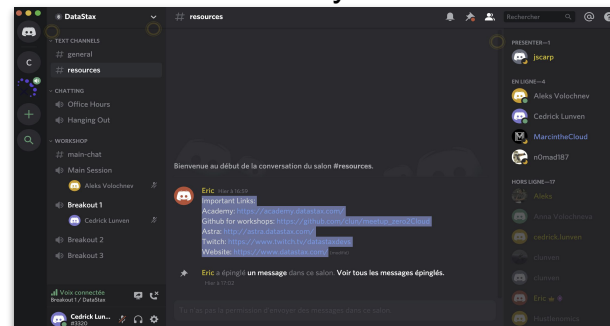
Courses: youtube.com/DataStaxDevs



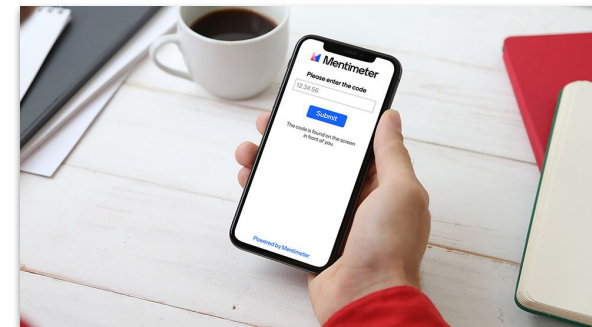
Runtime: dtsx.io/workshop



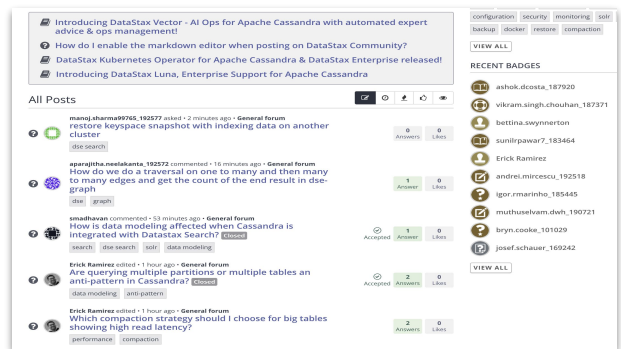
Questions: bit.ly/cassandra-workshop



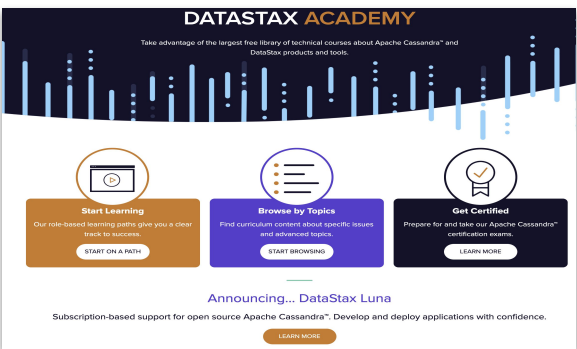
Quizz: menti.com



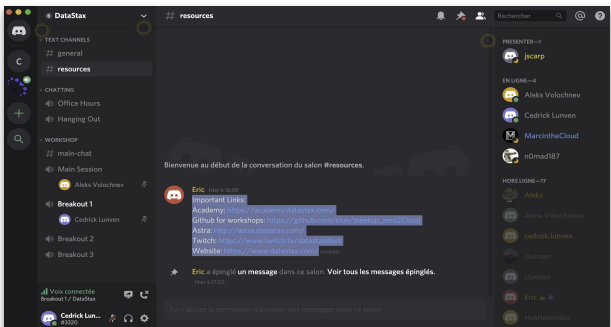
Forum: community.datastax.com



Training: academy.datastax.com



Chat: bit.ly/cassandra-workshop



Discord

DATASTAX
ASTRA

==



Fully managed Cassandra
Without the ops!

DataStax Astra



Global Scale

Put your data where you need it without compromising performance, availability, or accessibility.



No Operations

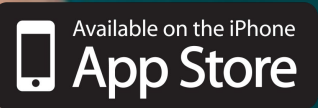
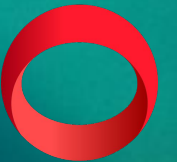
Eliminate the overhead to install, operate, and scale Cassandra.



5 Gig Free Tier

Launch a database in the cloud with a few clicks, no credit card required.

menti.com





<https://github.com/DataStax-Academy/workshop-storage-attached-indexes>



**What we will
cover:**

Storage Attached Indexes Workshop

- Indexing In Distributed Systems
- Secondary Indexing In Cassandra
- Storage Attached Indexes
- IoT Sensor Data Model Use Case
- Resources



**What we will
cover:**

Storage Attached Indexes Workshop

- Indexing In Distributed Systems
- Secondary Indexing In Cassandra
- Storage Attached Indexes
- IoT Sensor Data Model Use Case
- Resources

Primary Indexes

```
CREATE TABLE users (  
    id            long,  
    firstName    text,  
    lastName     text,  
    country      text,  
    created      timestamp,  
    PRIMARY KEY (id)  
);
```

Primary Index or Partition Key



```
SELECT firstName, lastName  
FROM users  
WHERE id = 100;
```

Single Server Database



Distributed Database



Non-Primary Key Queries

```
CREATE TABLE users (  
    id            long,  
    firstName    text,  
    lastName     text,  
    country      text,  
    created      timestamp,  
    PRIMARY KEY (id)  
);
```

```
SELECT firstName, lastName  
FROM users  
WHERE country = 'FR';
```

Single Server Database

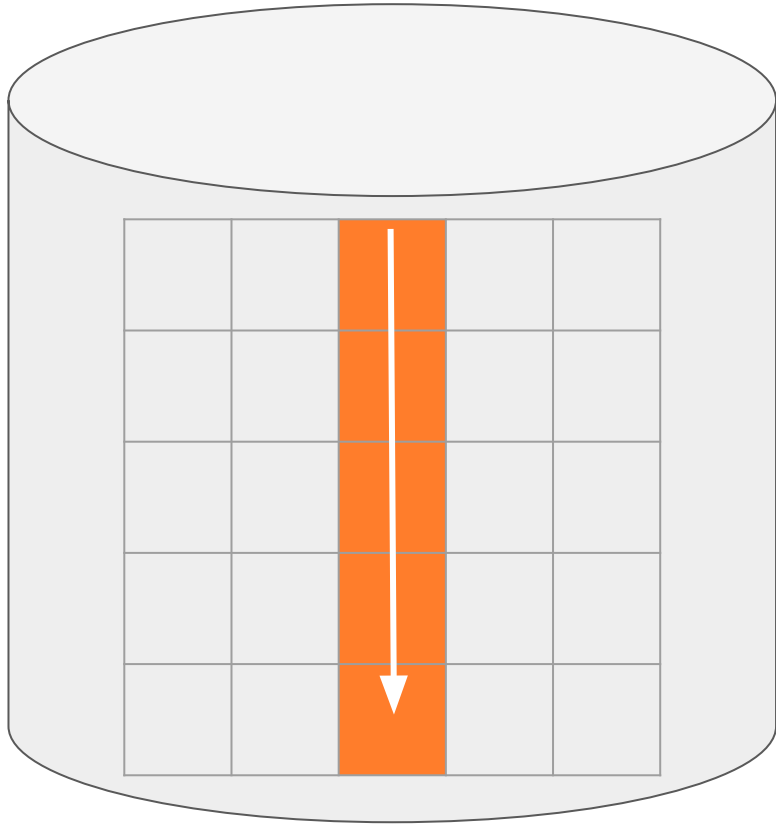


Distributed Database

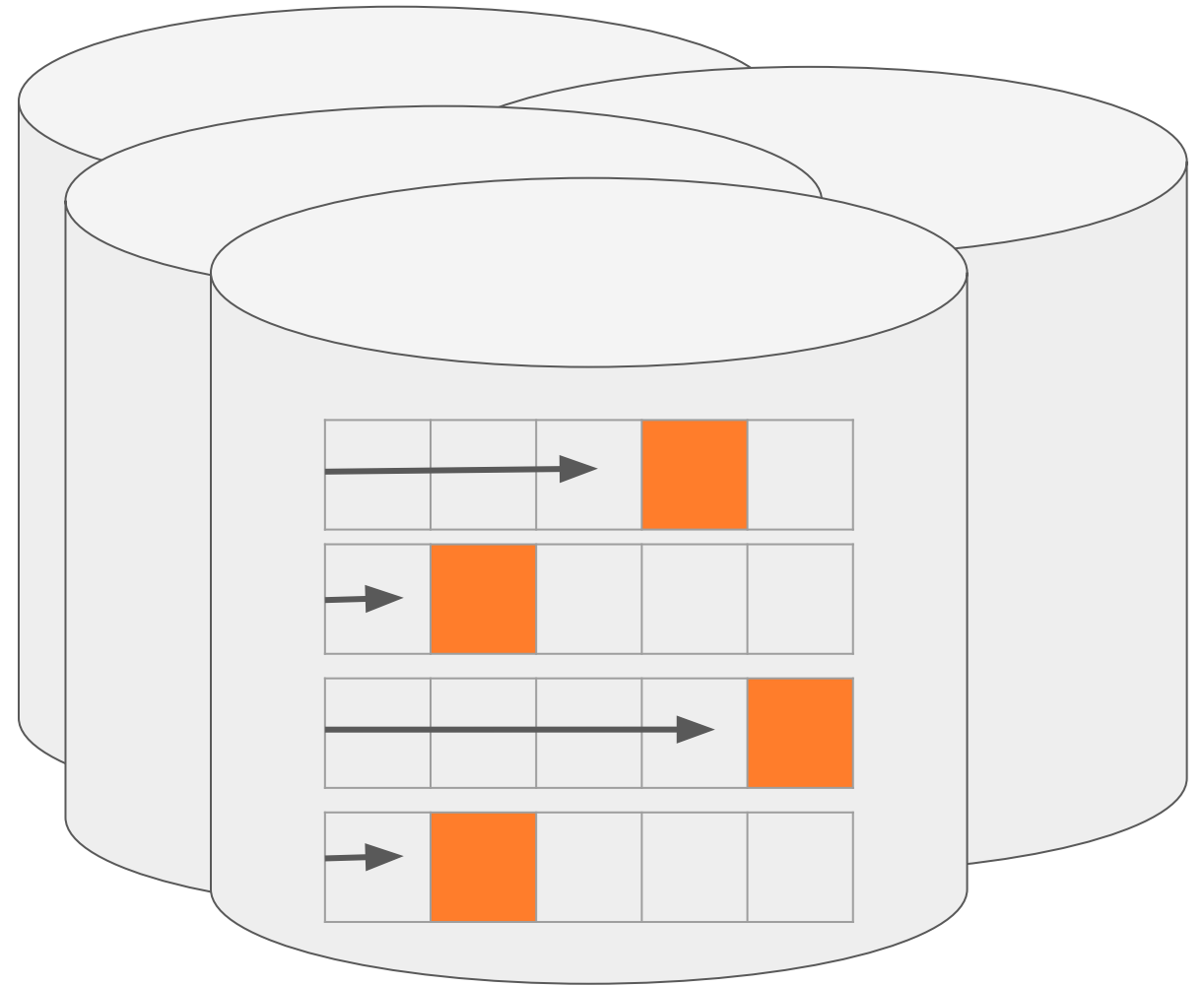




Why?



Relational Systems



Distributed Systems



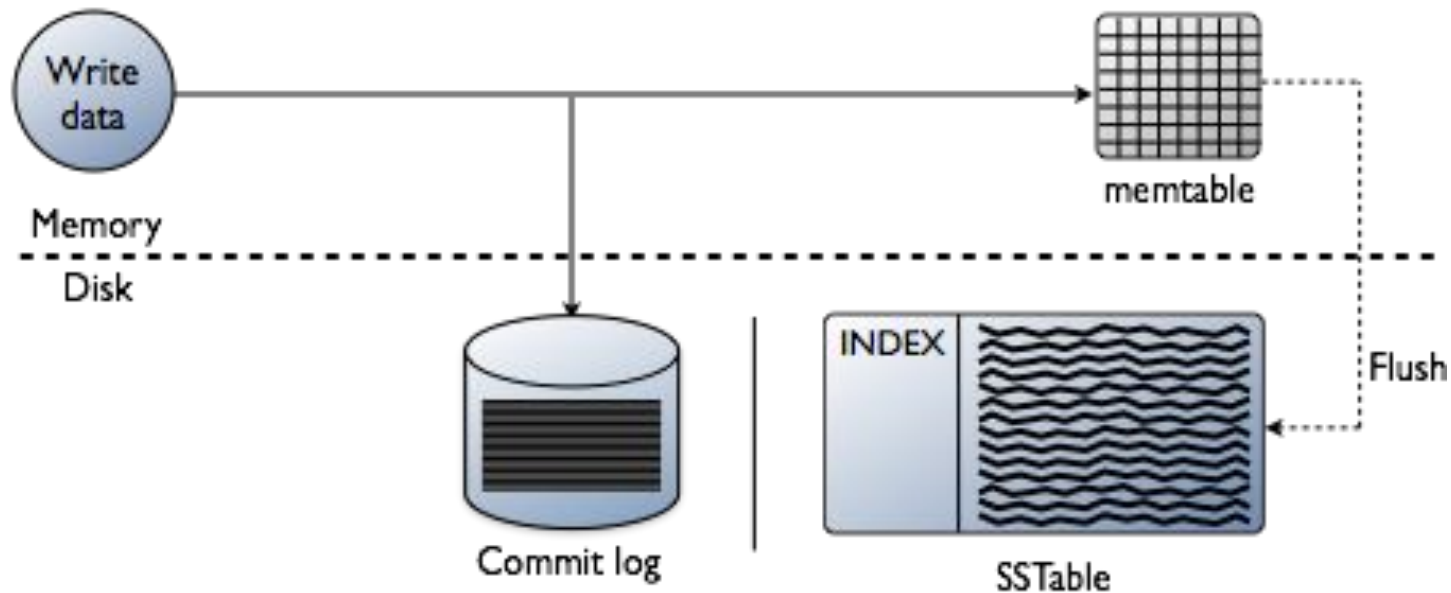
**What we will
cover:**

Storage Attached Indexes Workshop

- Indexing In Distributed Systems
- Secondary Indexing In Cassandra
- Storage Attached Indexes
- IoT Sensor Data Model Use Case
- Resources

In the Beginning...

Secondary Index (aka 2i)



Pros

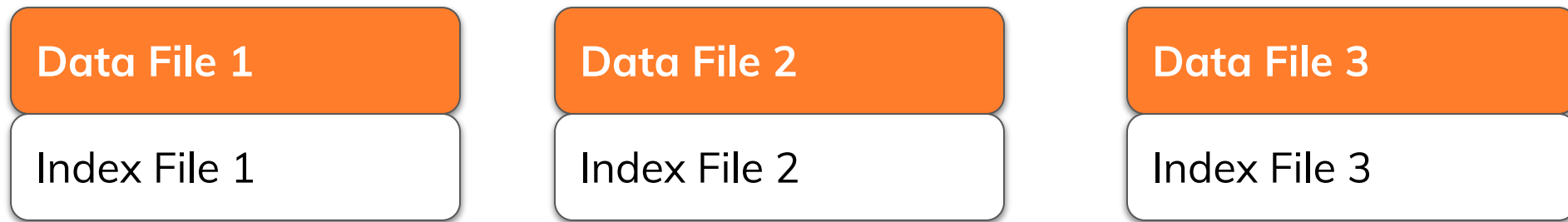
- Arbitrary columns in where clause
- Inequality ranges
- Disk space usage linear

Cons

- Reindex on each write
- Difficult for operators to manage

SASI

Storage Attached Secondary Index



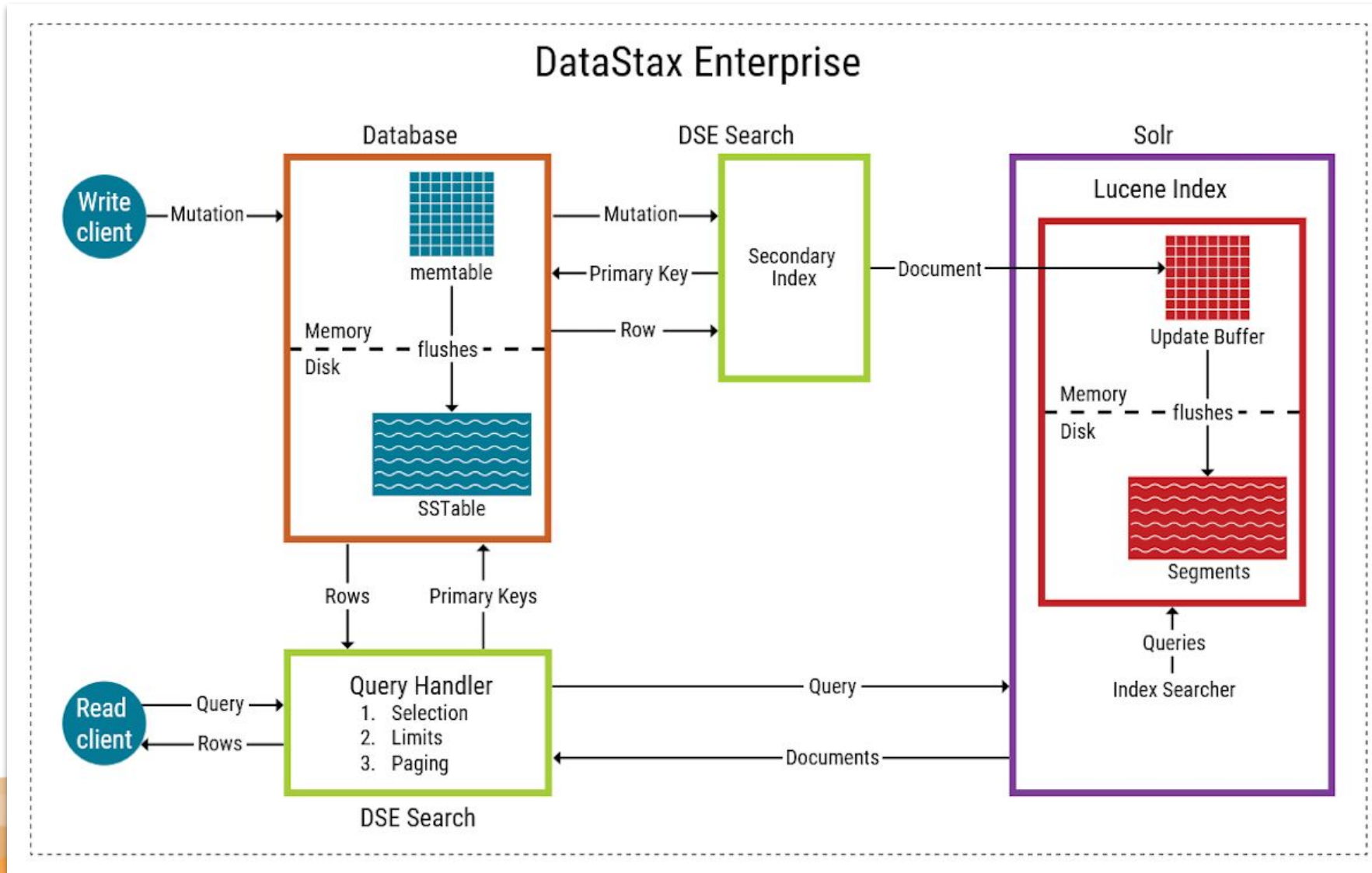
Pros

- Tokenized text search
- Fast range scans
- In-memory Indexing

Cons

- Huge index files
- Difficult for users to understand

DSE Search



Pros

- Full text search
- Efficient Indexing
- In-memory Indexing

Cons

- Full Solr engine on top of Cassandra
- Integration between CQL and Solr



**What we will
cover:**

Storage Attached Indexes Workshop

- Indexing In Distributed Systems
- Secondary Indexing In Cassandra
- Storage Attached Indexes
- IoT Sensor Data Model Use Case
- Resources

DDL

```
CREATE CUSTOM INDEX ON [keyspace_name.]table_name (column_name) USING  
'StorageAttachedIndex' WITH OPTIONS = {'case_sensitive': true/false, 'normalize':  
true/false};
```

Developers Can Write Flexible Queries

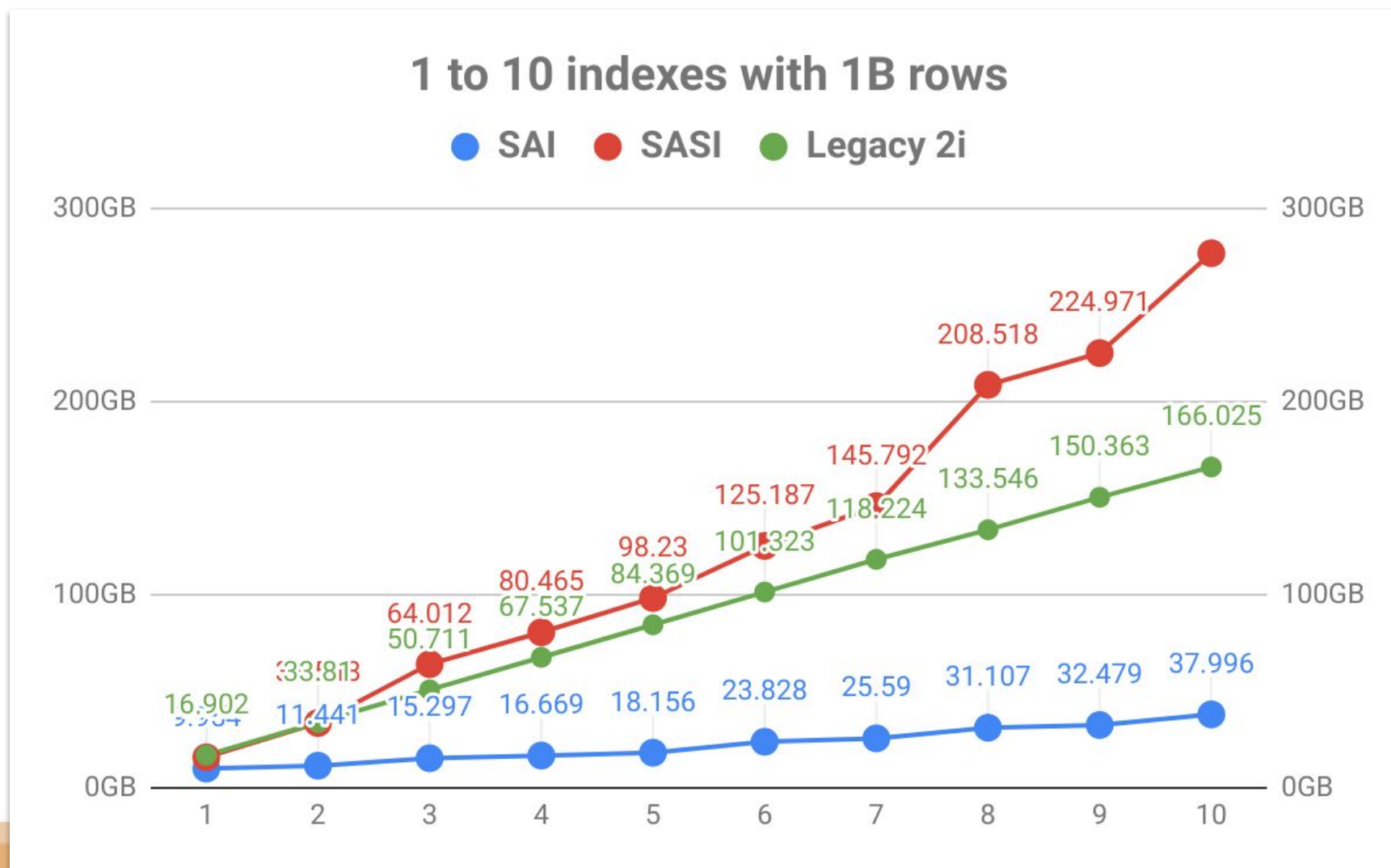
```
CREATE TABLE FOO ( id UUID PRIMARY KEY, c1 int, c2 text, c3 text);
```

SAI makes this possible in Cassandra - **SELECT * FROM FOO WHERE c2 = 'A' AND c3 = 'B'**

Pain points solved:

- Code around restrictive C* read path
- Duplicating denormalized data to query non PK fields
- The painful process of learning about all the edge cases with 2is
- Having to wait for DSE Search or other Search tools to be integrated with C*

Density Expectations



Coming to Apache Cassandra!

CEP-7: Storage Attached Index

- Being added post-4.0
- Ongoing conversations: **dev@cassandra.apache.org**



How does this change my
Cassandra data model?

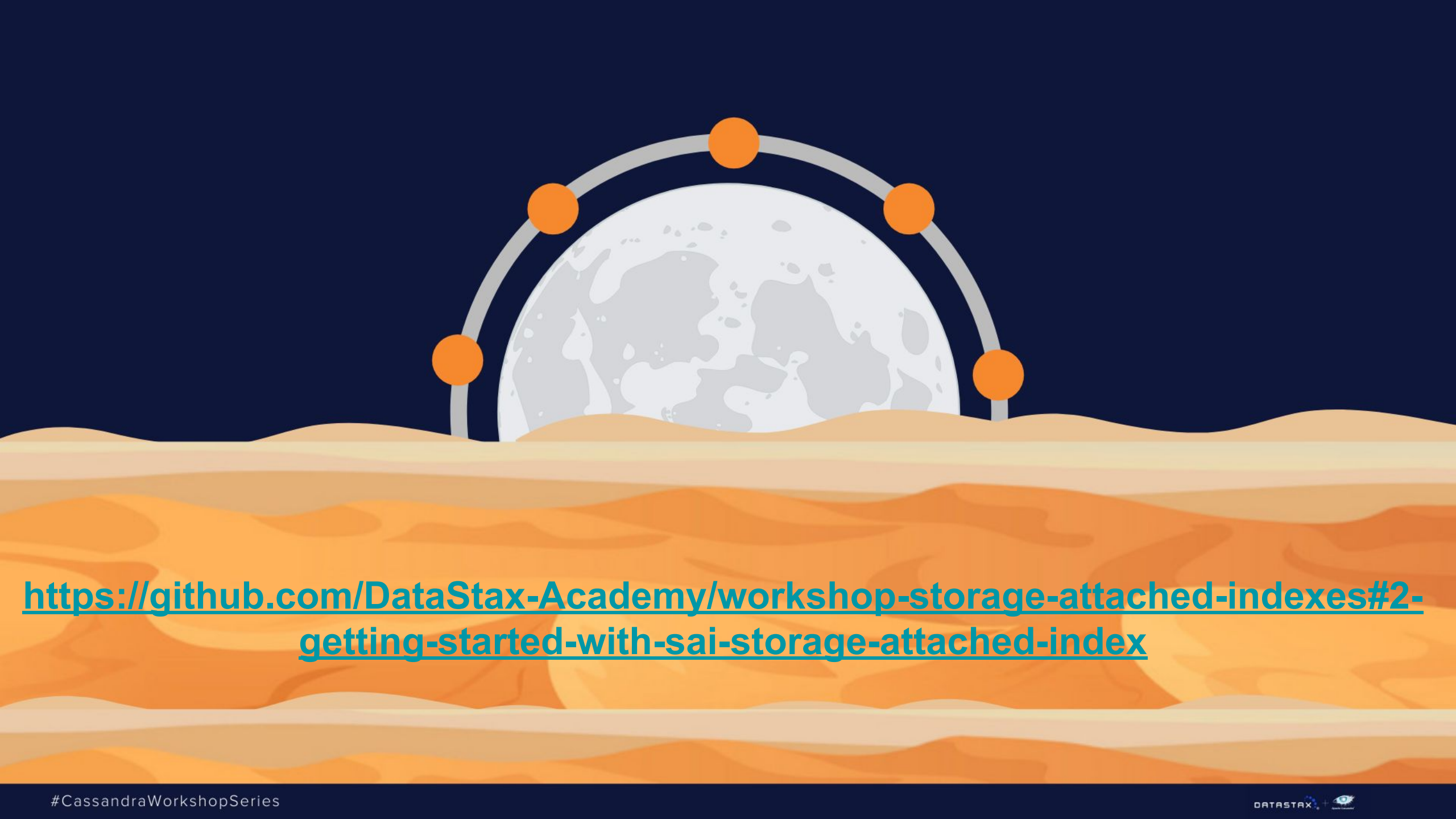


For new data models

**Index columns before
denormalization**

**Start with a single table and
build indexes from there**

**denormalize for SLA's within single digit millis*



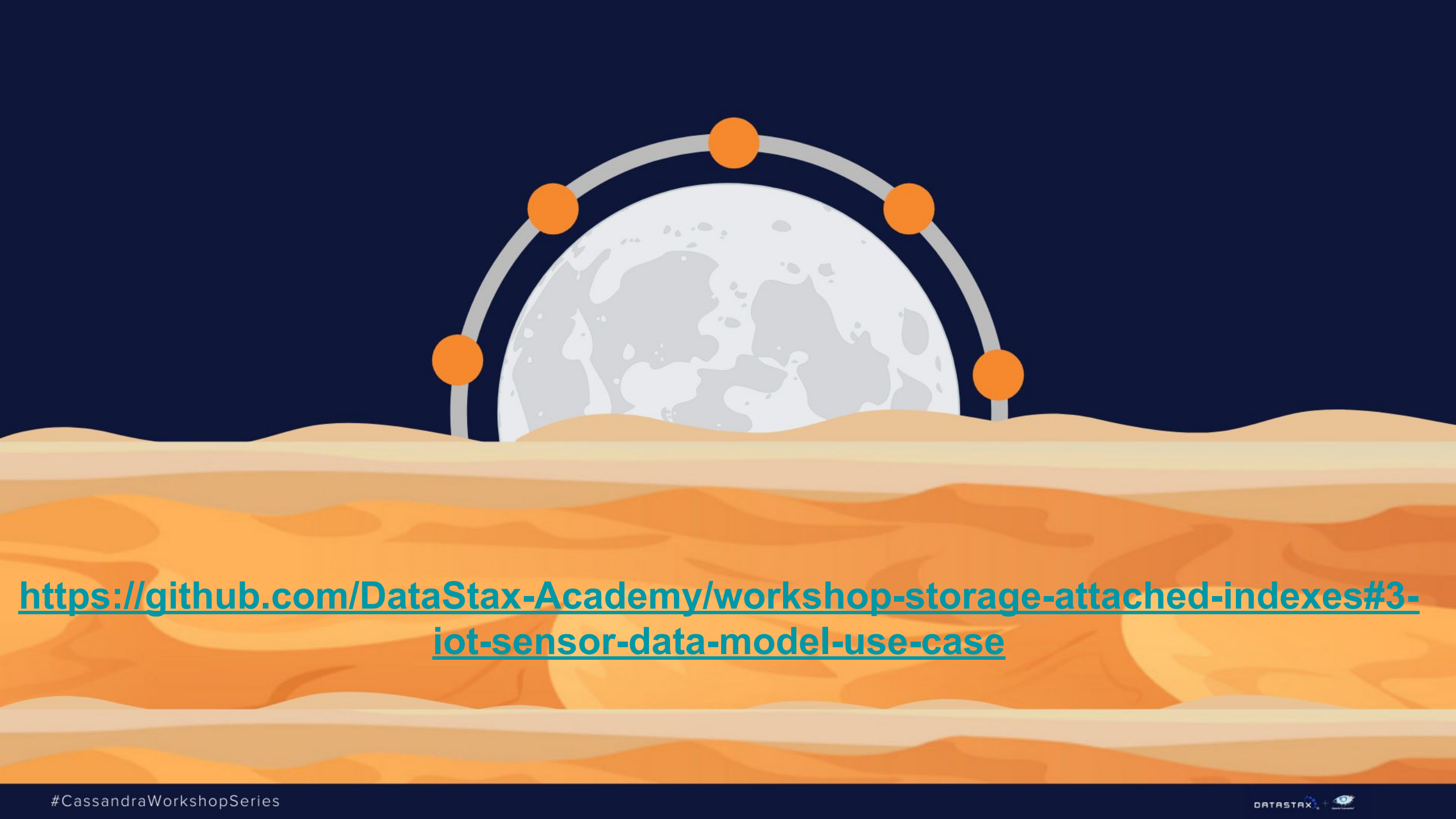
<https://github.com/DataStax-Academy/workshop-storage-attached-indexes#2-getting-started-with-sai-storage-attached-index>



**What we will
cover:**

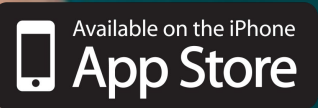
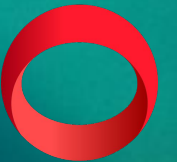
Storage Attached Indexes Workshop

- Indexing In Distributed Systems
- Secondary Indexing In Cassandra
- Storage Attached Indexes
- IoT Sensor Data Model Use Case
- Resources



[https://github.com/DataStax-Academy/workshop-storage-attached-indexes#3-
iot-sensor-data-model-use-case](https://github.com/DataStax-Academy/workshop-storage-attached-indexes#3-iot-sensor-data-model-use-case)

menti.com





**What we will
cover:**

Storage Attached Indexes Workshop

- Indexing In Distributed Systems
- Secondary Indexing In Cassandra
- Storage Attached Indexes
- IoT Sensor Data Model Use Case
- Resources

MORE LEARNING!!!!

Developer site: datastax.com/dev

- Developer Stories
- New hands-on learning scenarios with Katacoda
 - Try it Out
 - Cassandra Fundamentals
 - New Data Modeling course in progress
 - <https://katacoda.com/datastax>

Classic courses available at [DataStax Academy](https://datastax.com/academy)



Katacode

Developer Resources

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

We are on Youtube - Twitter - Twitch!

MATERIALS

Slides and exercises for this course are available at

<https://github.com/DataStax-Academy/workshop-storage-attached-indexes>



Thank You

