

DataStax

Developers

Build a clone of Good Reads
with Spring Framework

An Application built by

javabrains.io



Director of Developer Relations



- Trainer
- Public Speaker
- Developers Support
- Developer Applications
- Developer Tooling

- Creator of ff4j (ff4j.org)
- Maintainer for 8 years+

- Happy developer for 14 years
- Spring Petclinic Reactive & Starters
- Implementing APIs for 8 years



Cédrick Lunven



Developer Advocate Lead

dtsx.io/aleks



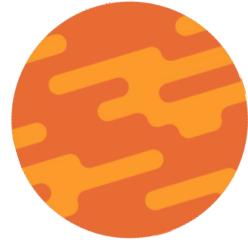
@aleks-volochnev
@HadesArchitect

- IT Exorcist
- Apache Cassandra™ certified
- Cloud Architect certified



Aleksandr Volochnev





S



Cedrick
Lunven



David
Dieruf



Rags
Srinivas



Artem
Chebotko



Stefano
Lottini



Aleksandr
Volochnev



Aaron
Ploetz



S



Jack
Fryer



Kirsten
Hunter



Gary
Harvey



Sonia
Siganporia



Ryan
Welford



David
Gilardi



DataStax Developers Crew

01



Housekeeping
Live and Hands-on

02

System Design
Use case and Technologies

03

Database Setup
Astra, Schema, Data

04

Connect & Use Data
SDK, Spring Data

05

Web and Security
Controller, View, SSO

06

What's next?
Quiz, Homework, Agenda



01



Housekeeping Live and Hands-on

02

System Design Use case and Technologies

03

Database Setup Astra, Schema, Data

04

Connect & Use Data SDK, Spring Data

05

Web and Security Controller, View, SSO

06

What's next? Quiz, Homework, Agenda

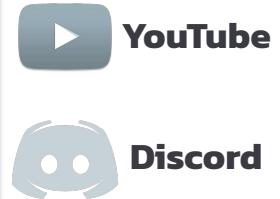
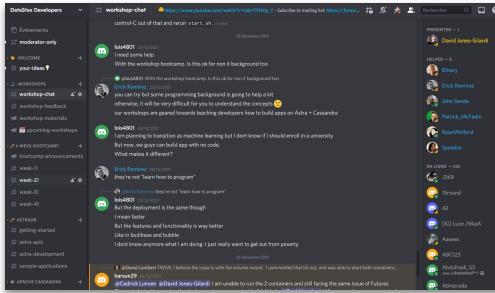


Agenda

Livestream: youtube.com/DataStaxDevs

Questions: https://dtsx.io/discord

Agenda



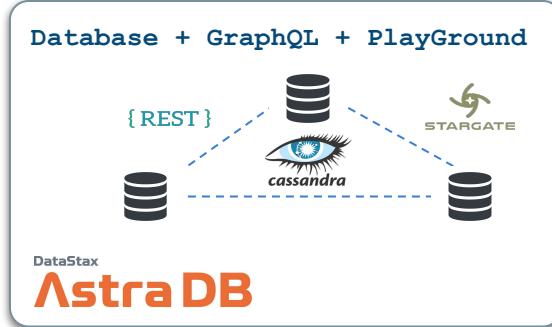
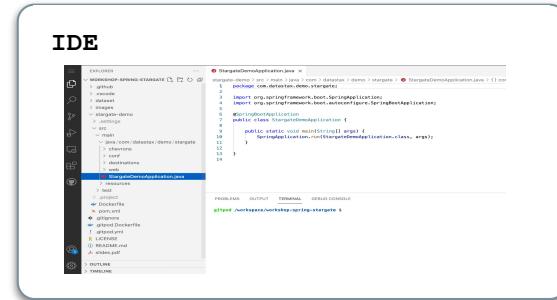
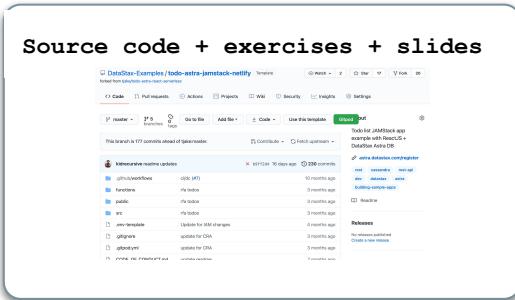
Games and quizzes: menti.com

How much experience do you have with the Spring Framework ?

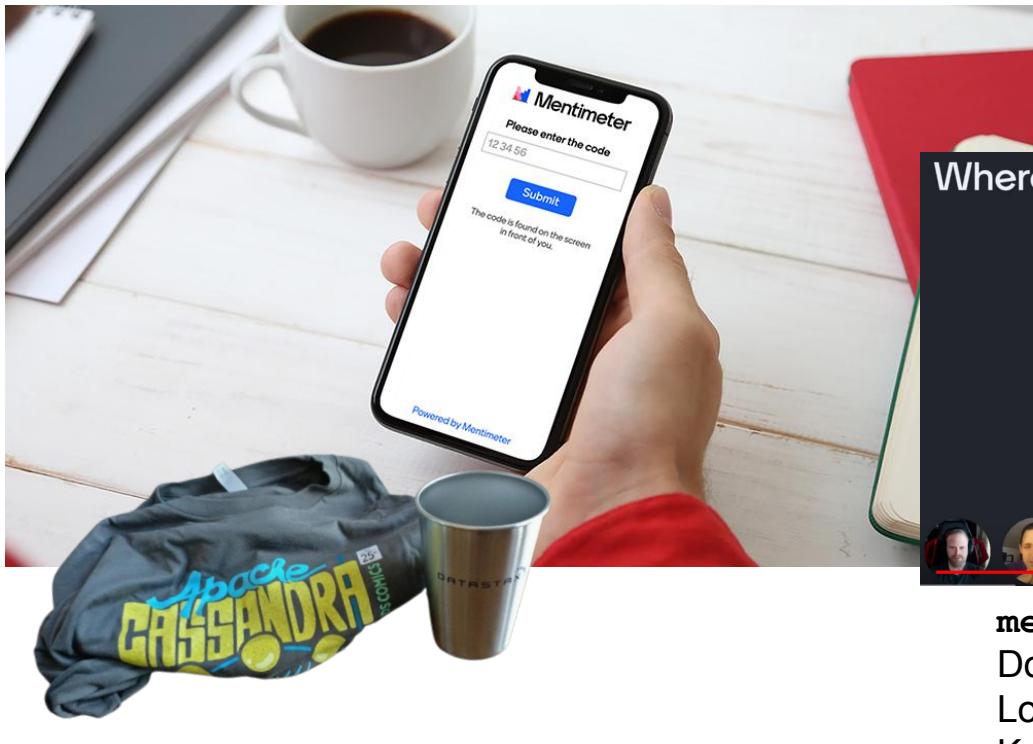


Live Sessions

Nothing to install !



Hands-On Housekeeping



Where are you from?



menti.com ⇒ enter code
Don't answer in YT chat
Look at phone (not at YT)
Keep it open for later

"Menti" for survey and quiz !

01



Housekeeping
Live and Hands-on

02

System Design
Use case and Technologies

03

Database Setup
Astra, Schema, Data

04

Connect & Use Data
SDK, Spring Data

05

Web and Security
Controller, View, SSO

06

What's next?
Quiz, Homework, Agenda



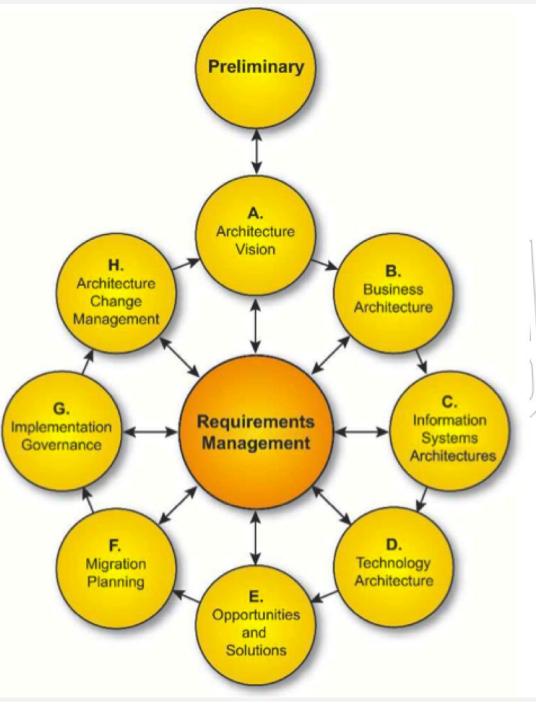
Agenda

Discover books
you'll love

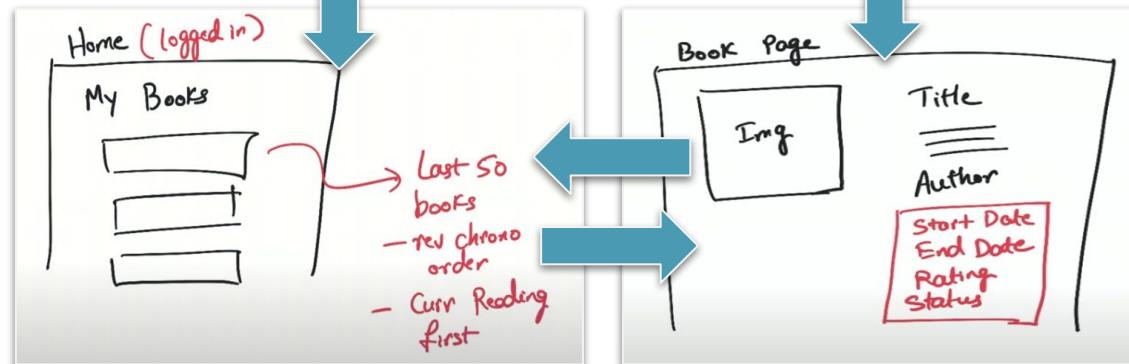
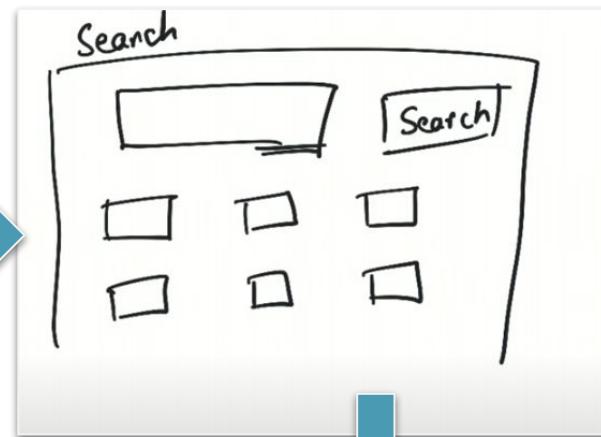
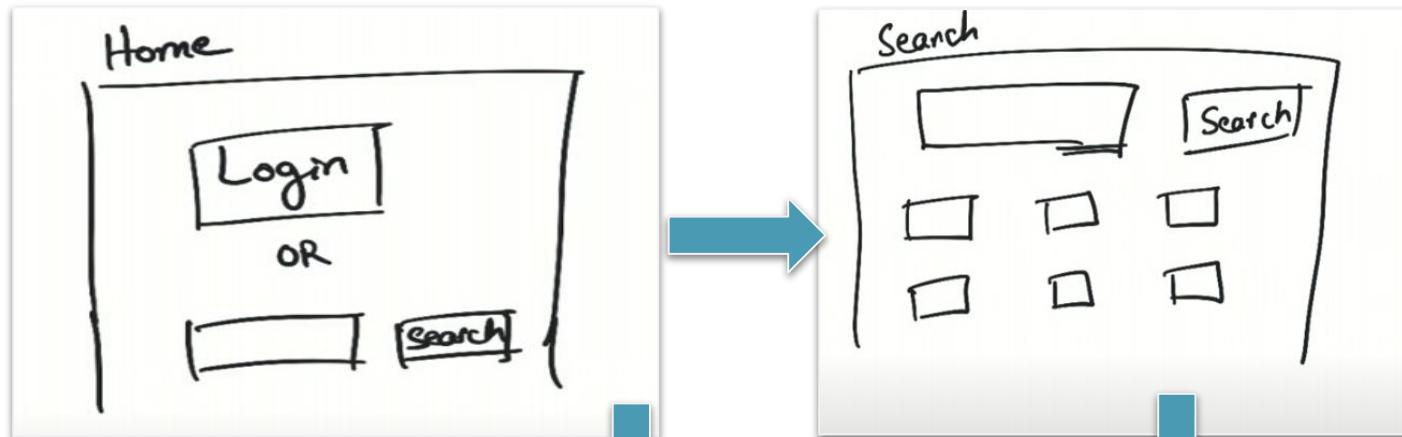


Find, get, and share books you love
with **everyone in the world**, in
realtime and all in
a single application.

(...a *clone of Good Reads*)

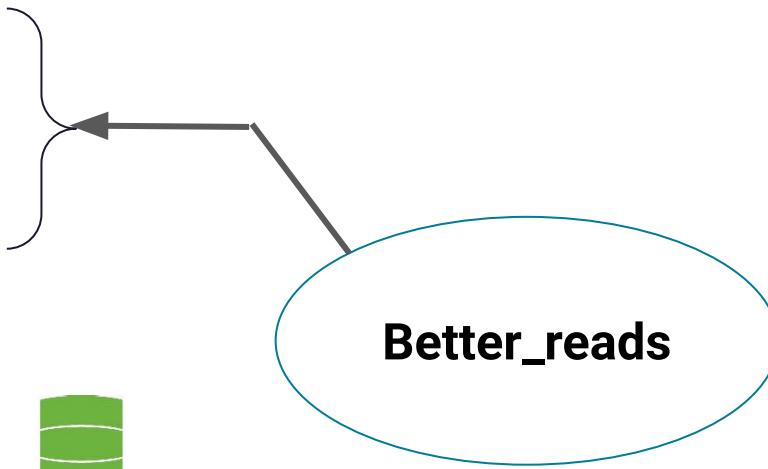


A. Architecture Vision



B. Business Architecture

- *Web app*
- *Focus on Backend*
- *Single Language*
- *Known Framework*

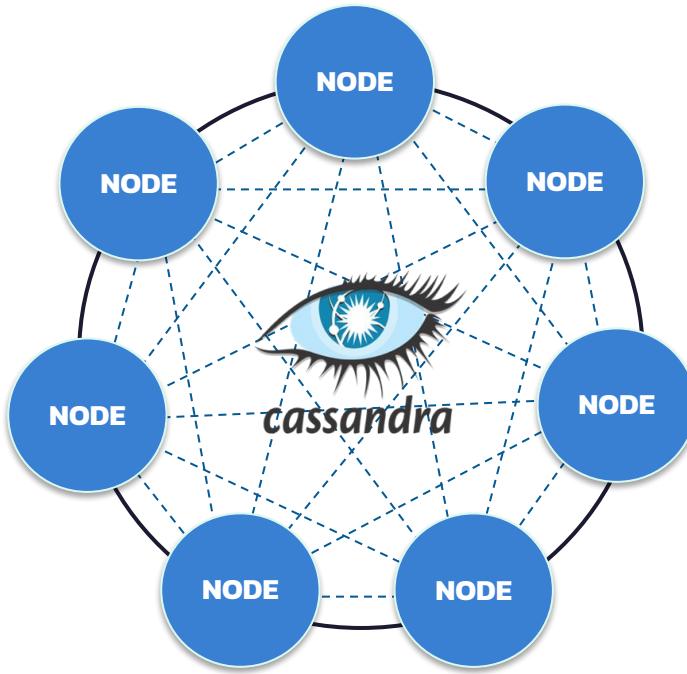


 *Thymeleaf*



C. Information System





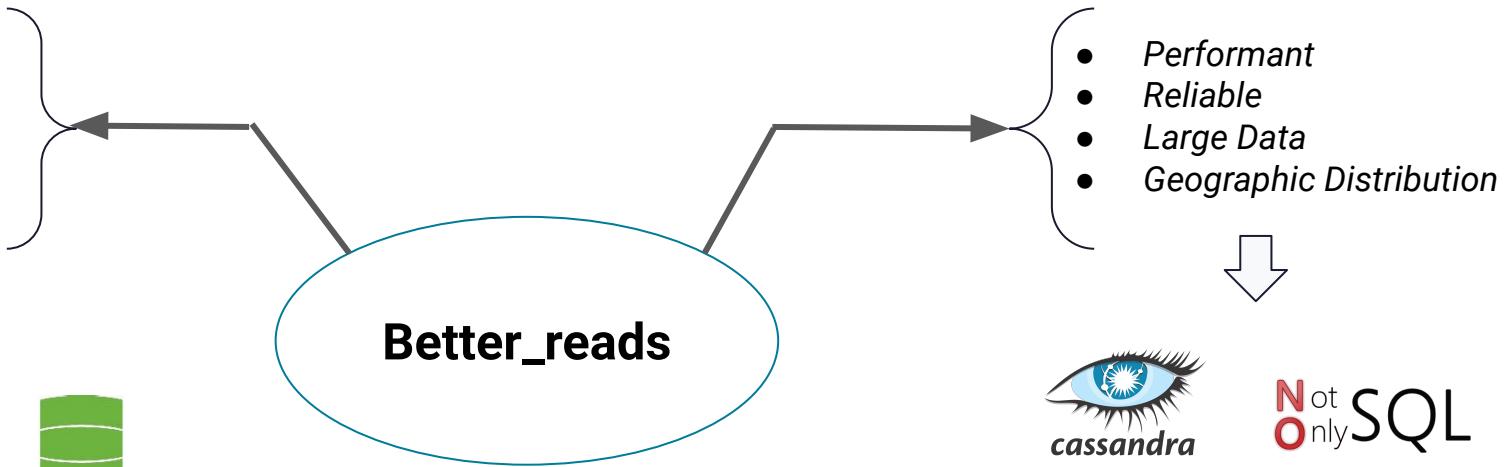
- Big Data Ready
- Read / Write Performance
- Linear Scalability
- Highest Availability
- Self-Healing and Automation
- Geographical Distribution
- Platform Agnostic
- Vendor Independent

Introduction to Apache Cassandra™

- Web app
- Focus on Backend
- Single Language
- Known Framework



 Thymeleaf



Not
Only SQL



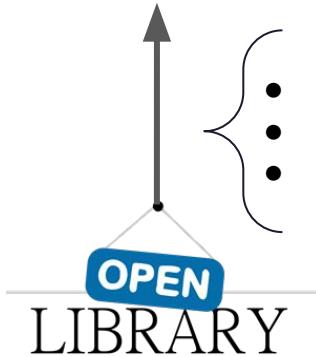
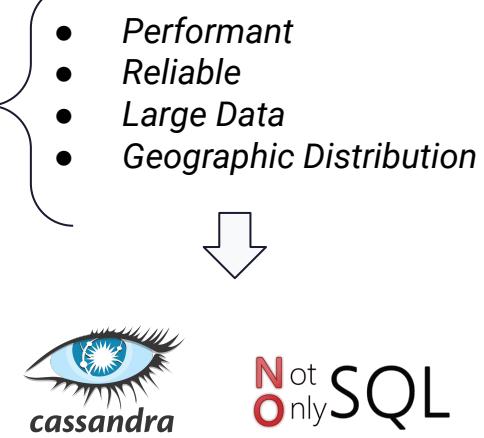
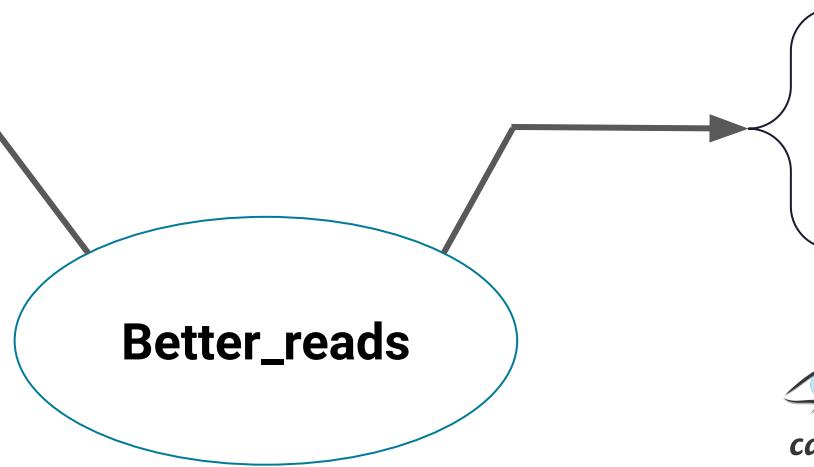
C. Information System



- Web app
- Focus on Backend
- Single Language
- Known Framework



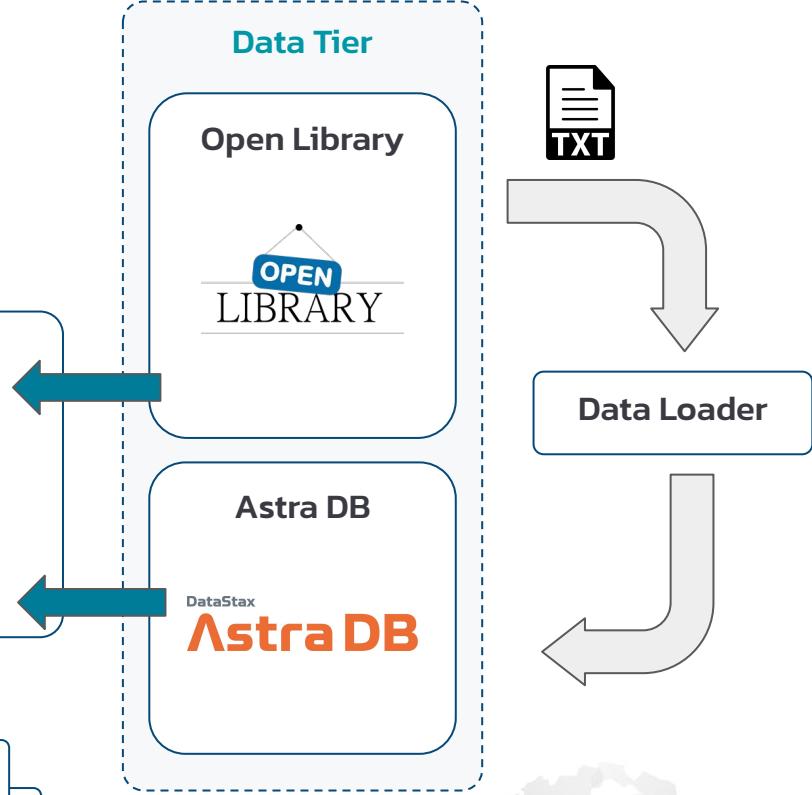
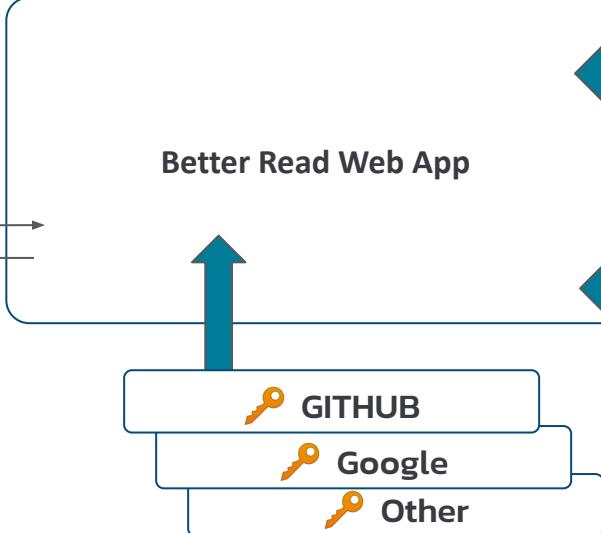
 Thymeleaf



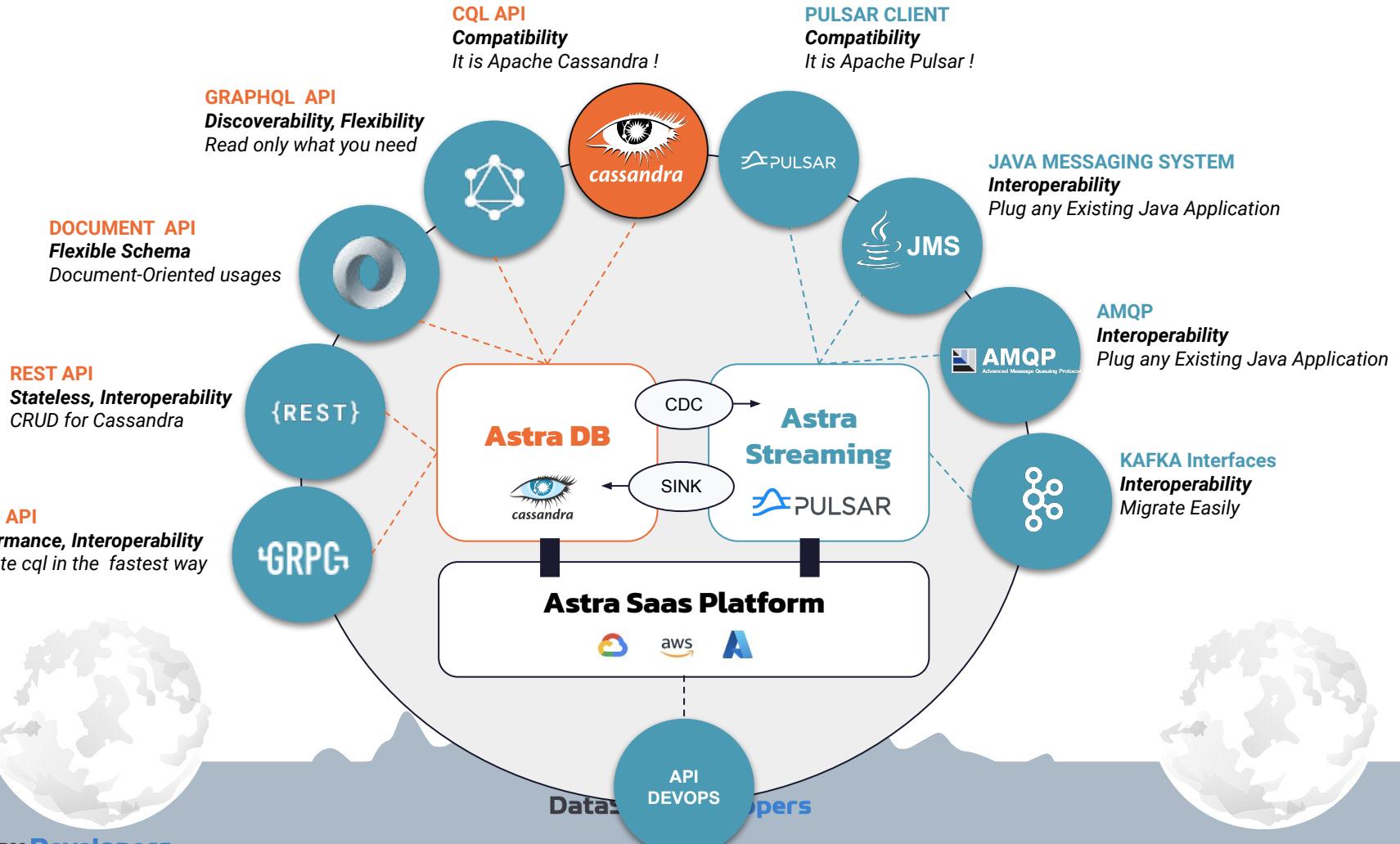
- DataSet (dump)
- Image
- Search API



C. Information System



D. Technology Architecture



01



Housekeeping
Live and Hands-on

02

System Design
Use case and Technologies

03

Database Setup
Astra, Schema, Data

04

Connect & Use Data
SDK, Spring Data

05

Web and Security
Controller, View, SSO

06

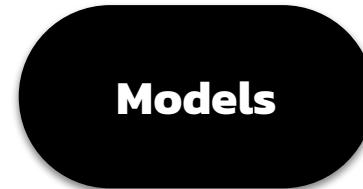
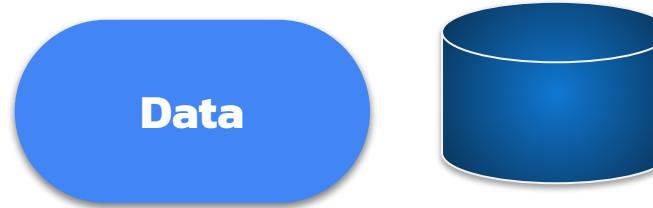
What's next?
Quiz, Homework, Agenda



Data Modelling



1. Analyze raw data
2. Identify entities, their properties and relations
3. Design tables, using **normalization** and foreign keys.
4. Use JOIN when doing queries to join normalized data from multiple tables



The diagram shows two normalized tables: "Employees" and "Departments". A line connects the "departmentId" column in the "Employees" table to the "departmentId" column in the "Departments" table, indicating a foreign key relationship.

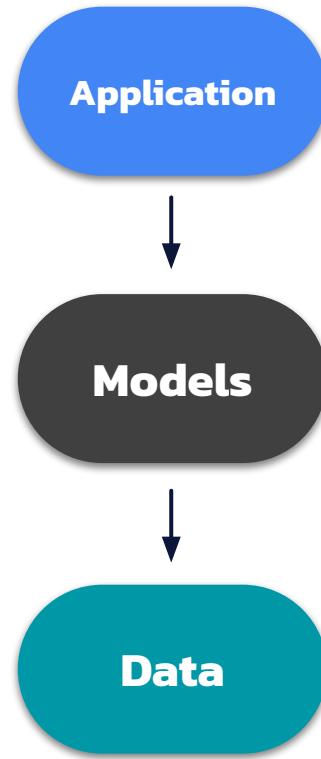
Employees		
userId	firstName	lastName
1	Edgar	Codd
2	Raymond	Boyce

Departments	
departmentId	department
1	Engineering
2	Math

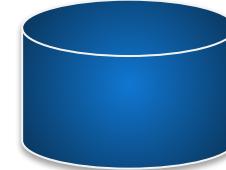


Relational Data Modelling

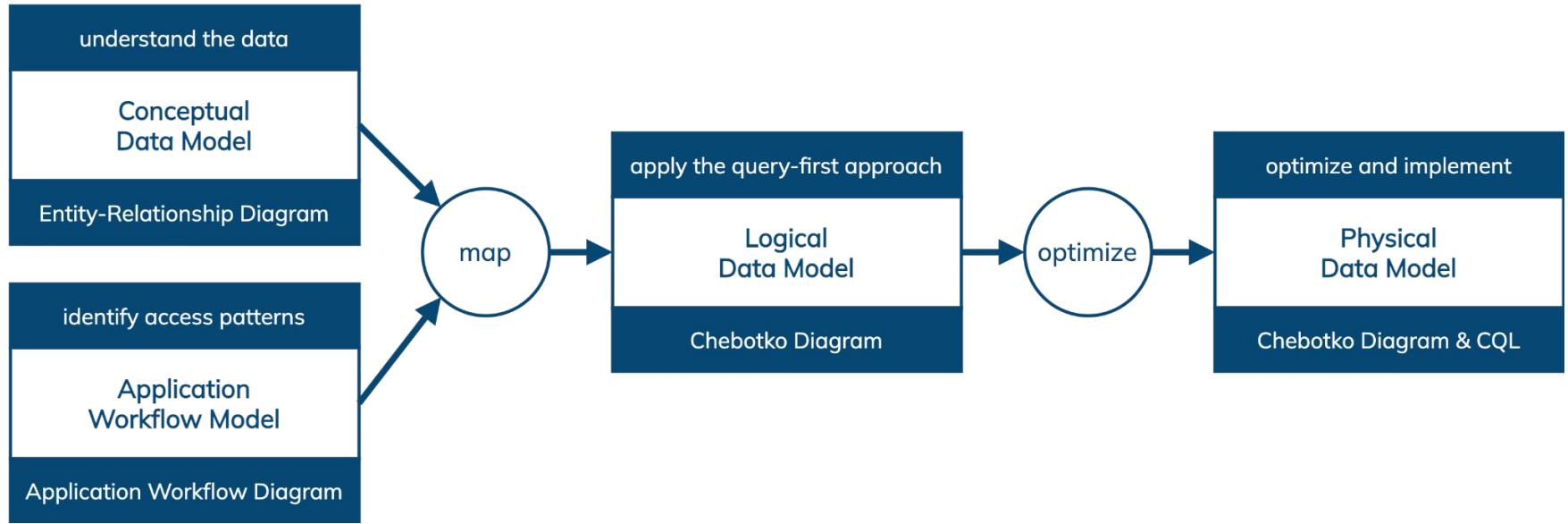
1. Analyze user behaviour
(customer first!)
2. Identify workflows, their dependencies
and needs
3. Define Queries to fulfill these workflows
4. Knowing the queries, design tables,
using **denormalization**.
5. Use BATCH when inserting or updating
denormalized data of multiple tables



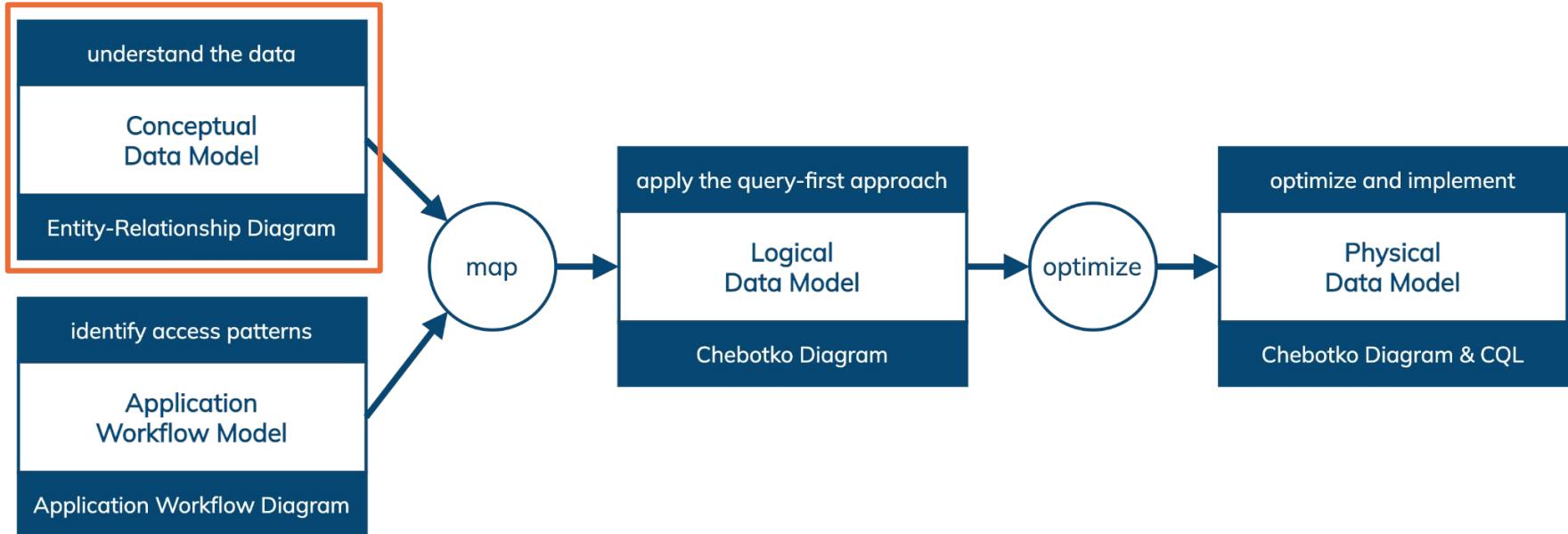
Employees			
userId	firstName	lastName	department
1	Edgar	Codd	Engineering
2	Raymond	Boyce	Math
3	Sage	Lahja	Math
4	Juniper	Jones	Botany



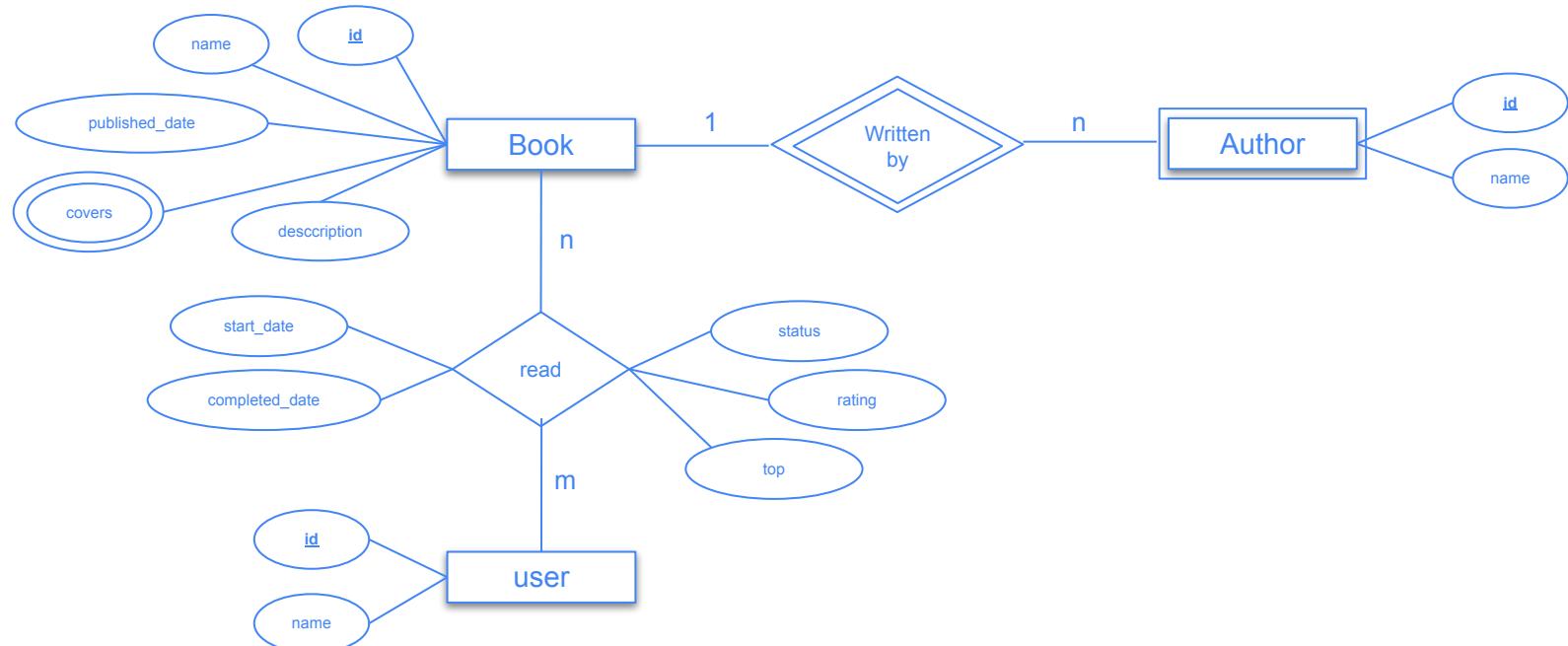
NoSQL Data Modelling



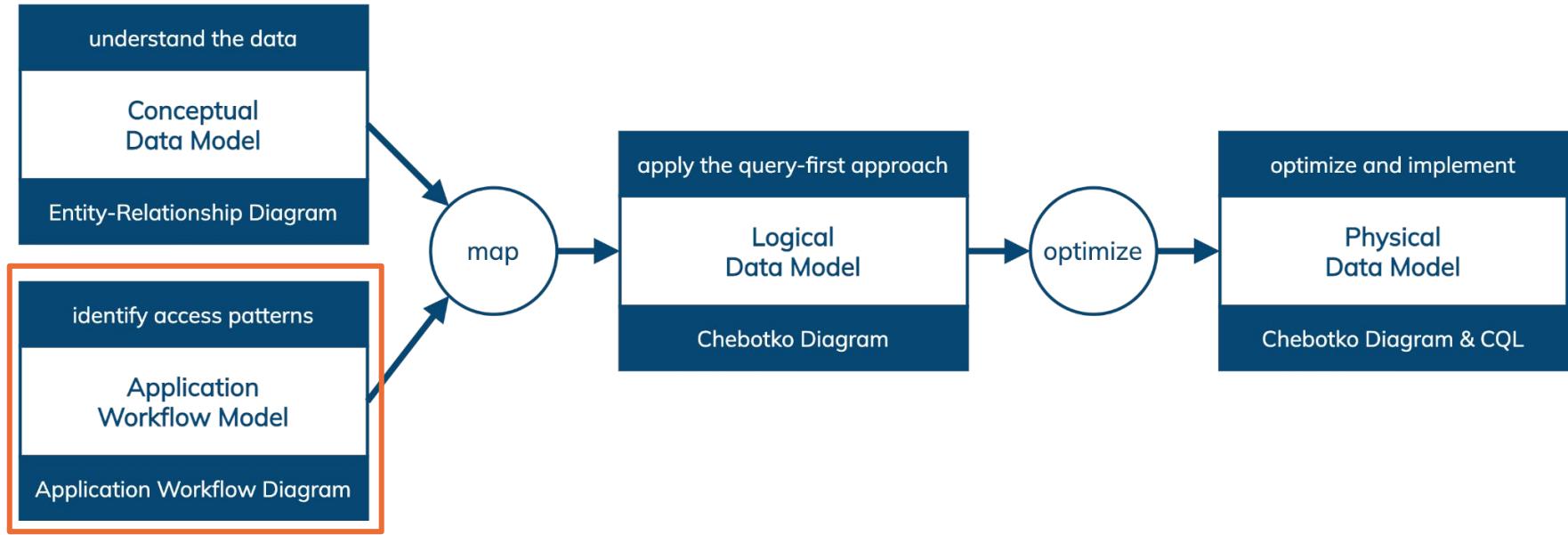
Data Modelling Methodology



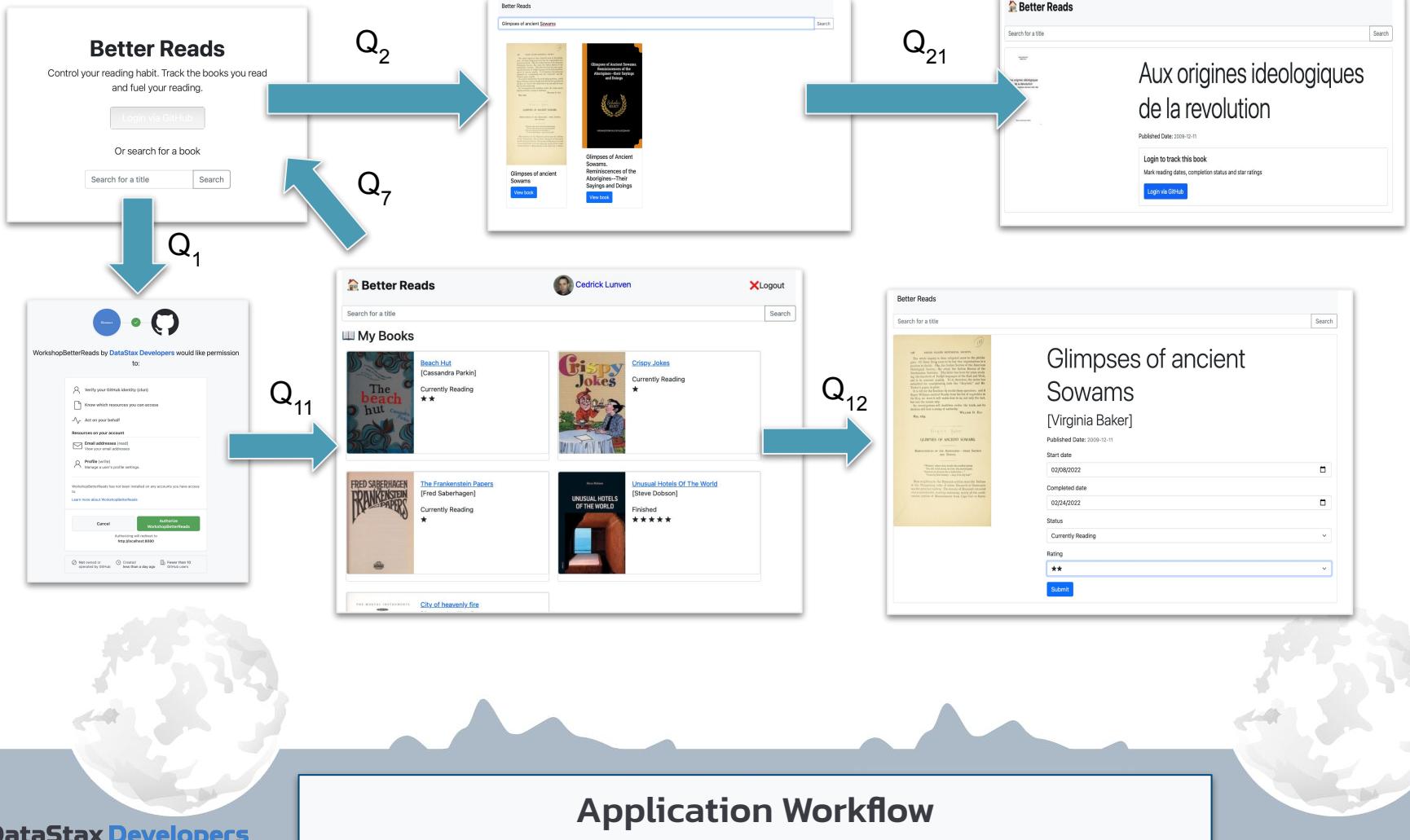
Data Modelling Methodology

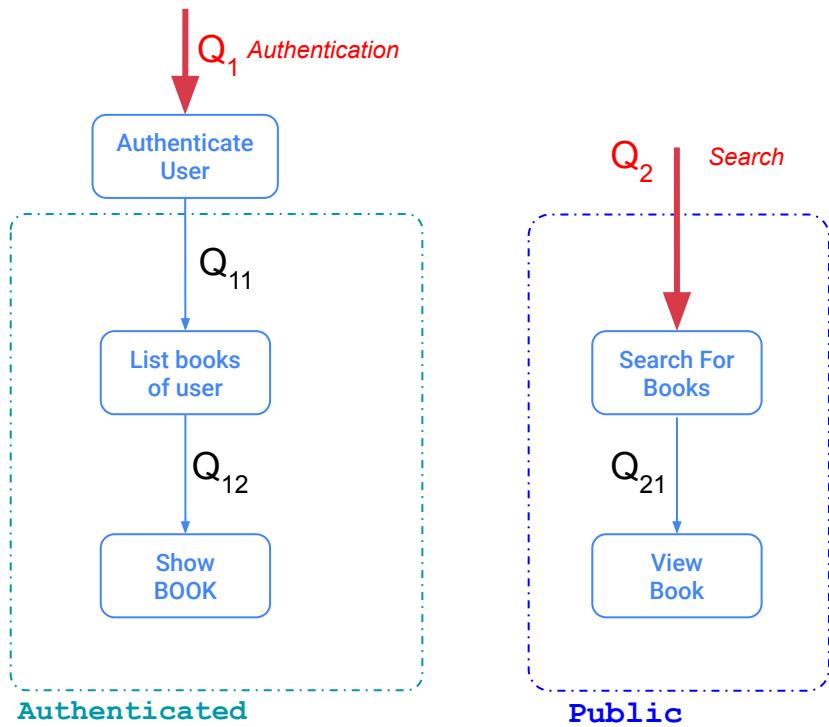


Entity-Relationship Diagram

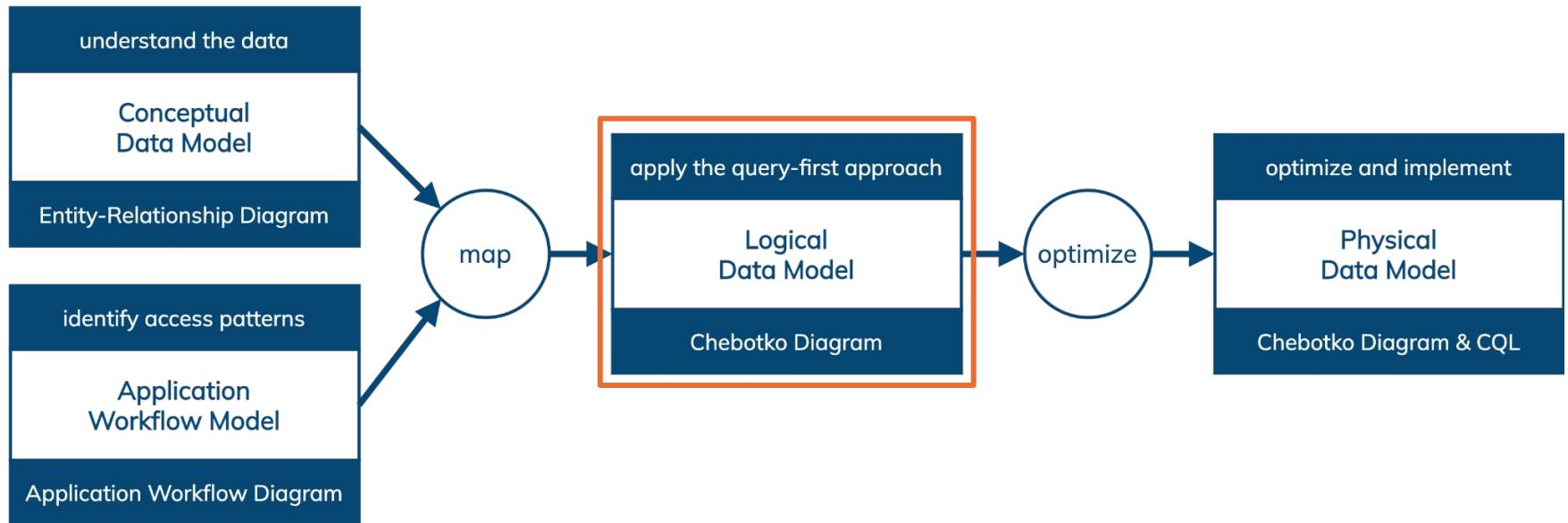


Data Modelling Methodology





Application Workflow



Data Modelling Methodology

Q₂₁

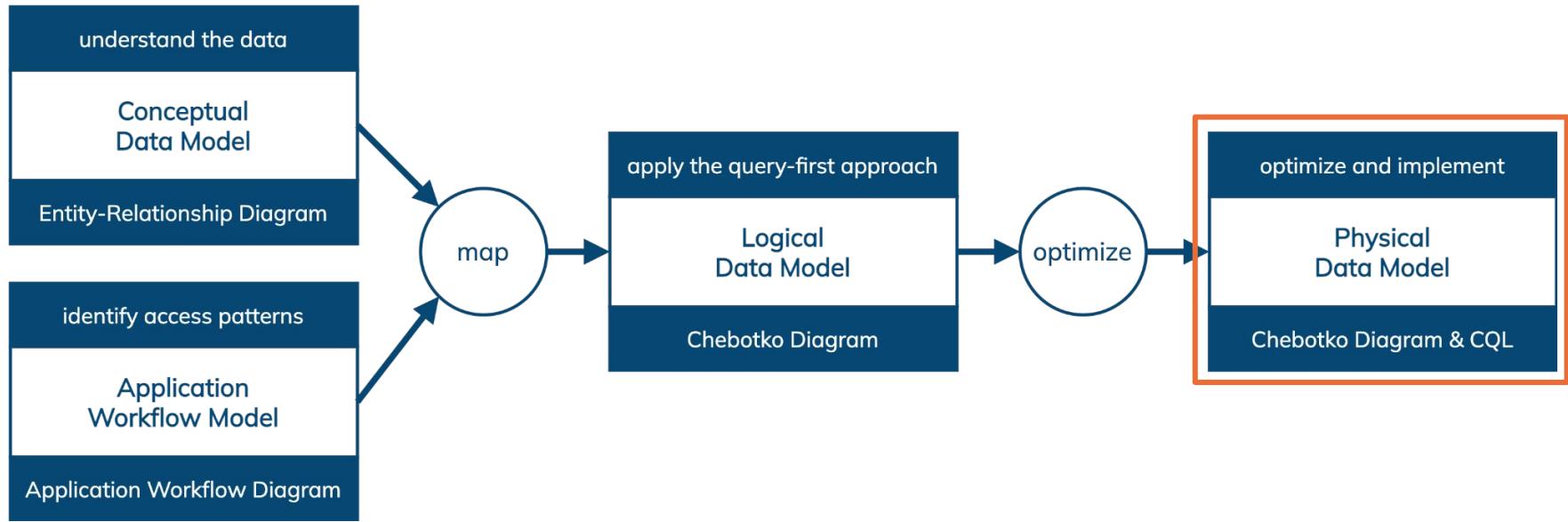
books_by_id	
id	K
author_ids	
author_names	
book_description	
book_name	
cover_ids	
published_date	

Q₁₂

book_by_user_and_bookid	
user_id	K
readingstatus	C↑
timeuuid	C↓
book_id	C↑
book_name	
author_names	
cover_ids	
rating	

Q₁₁

book_by_user	
user_id	K
book_id	K
rating	
reading_status	
started_date	
completed_date	



Data Modelling Methodology

Q₂₁

books_by_id

id	text	K
author_ids	list<text>	
author_names	list<text>	
book_description	text	
book_name	text	
cover_ids	list<text>	
published_date	date	

Q₁₂

book_by_user_and_bookid

user_id	text	K
readingstatus	text	
timeuuid	timeuuid	C↑
book_id	text	C↓
book_name	text	
author_names	list<text>	C↑
cover_ids	list<text>	
rating	int	

Q₁₁

book_by_user

user_id	text	K
book_id	text	K
rating	int	
reading_status	text	
started_date	date	
completed_date	date	



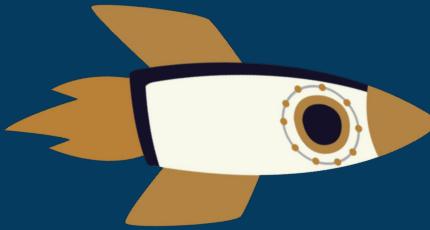
Physical Data Model



```
CREATE TABLE better_reads.book_by_id (
    id              text,
    author_id       list<text>,
    author_names    list<text>,
    book_description text,
    book_name       text,
    cover_ids       list<text>,
    published_date  date,
    PRIMARY KEY((id))
);

CREATE TABLE better_reads.books_by_user (
    user_id          text,
    book_id          text,
    book_name        text,
    author_names     list<text>,
    cover_ids        list<text>,
    readingstatus   text,
    timeuuid         timeuuid,
    rating           int,
    PRIMARY KEY ((user_id), readingstatus, timeuuid, book_id)
) WITH CLUSTERING ORDER BY (readingstatus ASC, timeuuid DESC, book_id ASC);
```

```
CREATE TABLE better_reads.book_by_user_and_bookid (
    user_id          text,
    book_id          text,
    rating           int,
    reading_status   text,
    started_date    date,
    completed_date  date,
    PRIMARY KEY ((user_id, book_id))
);
```



Hands-on (!github)

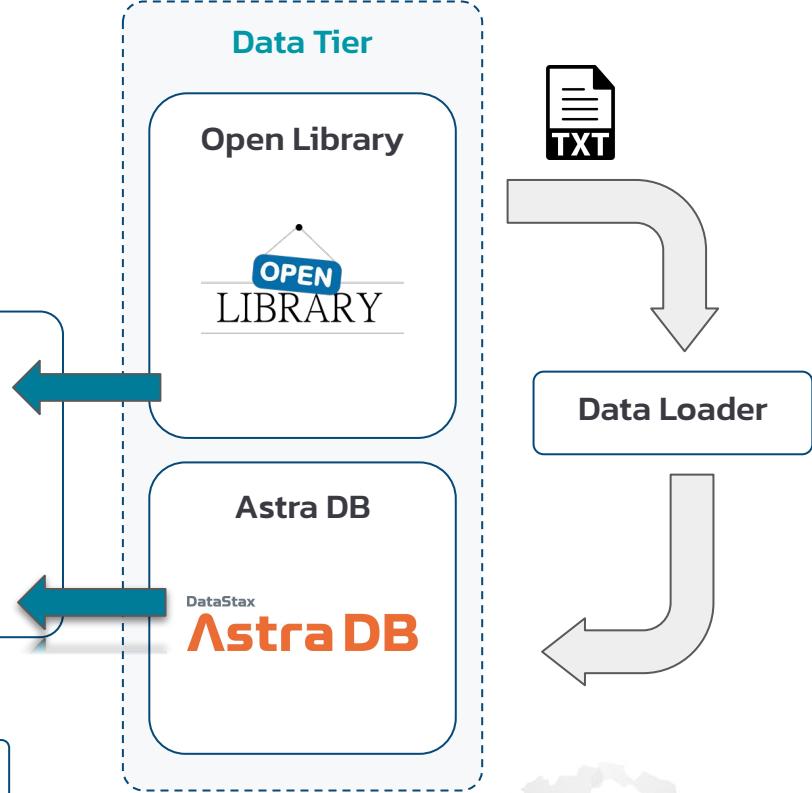
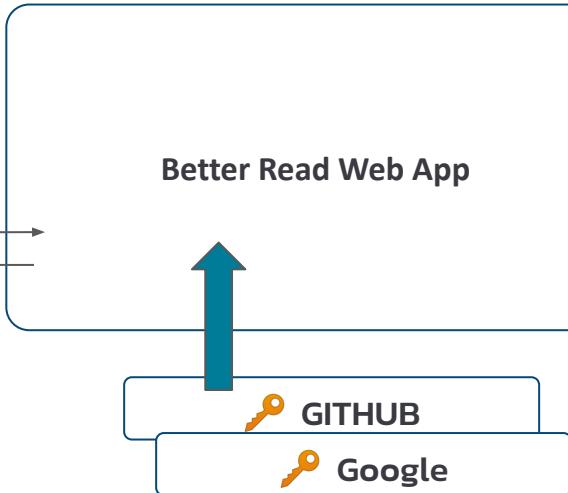
SETUP

- ✓ Create your Database (or resume)
- ✓ Create a token (if needed)



Data Loading





Welcome to Open Library



Read Free Library Books Online
Millions of books through Control Lending

Classic Books



Read

Developer Center

Developer Center (Home) | W

Welcome to the developer hub of Open Library. Here you can access data from, integrating with our API.

Accessing Data

Open Library offers free, public, low-latency examples of partners making use of our API. If you're a developer looking to search for specific books, their authors, or download our free data dumps.

Client Libraries

We maintain an [Official \(Python\) Client](#).

Contributing

OpenLibrary.org is an open source project. If you'd like to contribute, please see our [volunteer page](#). You can also explore our [partner resources](#).

Partners

Are you a library, publisher, or book distributor? See our [partner resources](#).

Developer Center

INTERN ARCHIVE

Open Library Data Dumps

Last edited by [ravbb](#)
November 17, 2021 | [History](#) Edit

Open Library provides dumps of all its data, generated every month. All of the dumps are formatted as tab separated files with the following columns:

- type – type of record (/type/edition, /type/work etc.)
- key – unique key of the record. (/books/OL1M etc.)
- revision – revision number of the record
- last_modified – last modified timestamp
- JSON – the complete record in JSON format

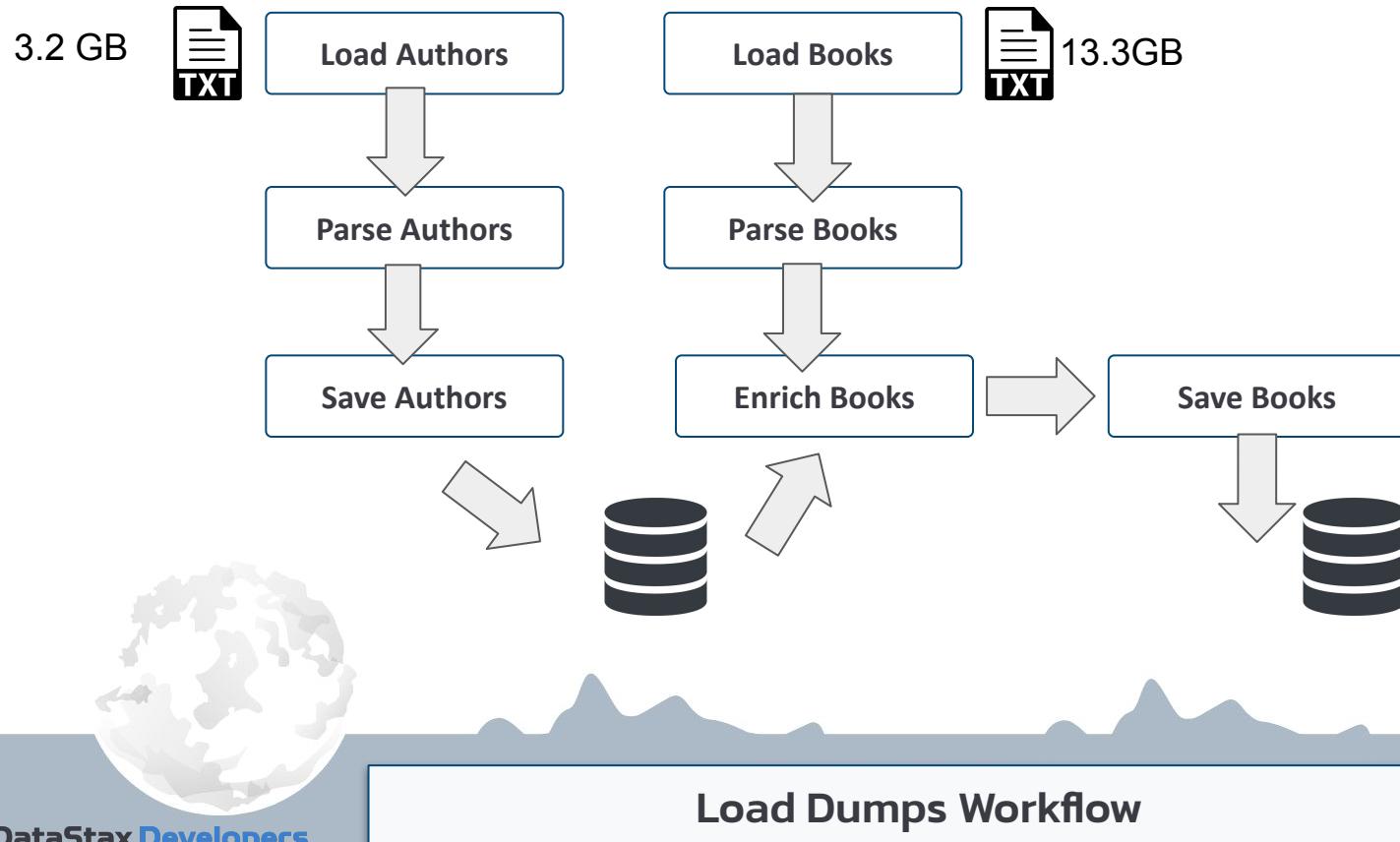
Dumps:

- [editions dump \(~ 7.0G\)](#)
- [works dump \(~ 2.0G\)](#)
- [authors dump \(~ 0.4G\)](#)
- [all types dump \(~ 10.0G\)](#): includes editions, works, authors, redirects, etc.
- [complete dump \(~ 23.3G\)](#): also includes past revisions of all the records in Open Library
- [ratings dump \(~ 5M\)](#): with columns: "Work Key, Edition Key (optional), Rating, Date"
- [reading log dump \(~ 20MB\)](#): with columns "Work Key, Edition Key (optional), Shelf Data"

books_by_id

id	K
author_ids	
author_names	
book_description	
book_name	
cover_ids	
published_date	

- Authors containing : Author id and Author Names
- Works containing the Books informations and only author_id



`mvn test -Dtest=io.javabrains.betterreads.Test02_ImportDataSetWithSpring`

```
22:14:14.577 INFO com.datastax.astra.sdk.AstraClient : [AstraClient] has been initialized.  
1: Saving author: Colombe Plante  
2: Saving author: Cabanes/  
3: Saving author: Damenie Louis  
4: Saving author: Damenie Louis  
5: Saving author: Jerome Maran  
6: Saving author: Alain Renk  
7: Saving author: Francois G. Ricaud  
8: Saving author: Jacqueline Bregetzer  
9: Saving author: Jacky Loufrani  
10: Saving author: Brigitte Engammare  
11: Saving author: Mouret
```

See you in 10 days... for the 9M authors.
See you in 280 days.. For the 260M books...



Load Dump Demo



Do not optimize too soon...

Should we do better ?

Can we do better ?

When is too soon?



Datastax Bulk



```
dsbulk unload \
  -url /workspace/workshop-betterreads/export/ \
  -k better_reads -t book_by_id -u token -p AstraCS:..... \
  -b /tmp/secure-connect-workshops.zip
```

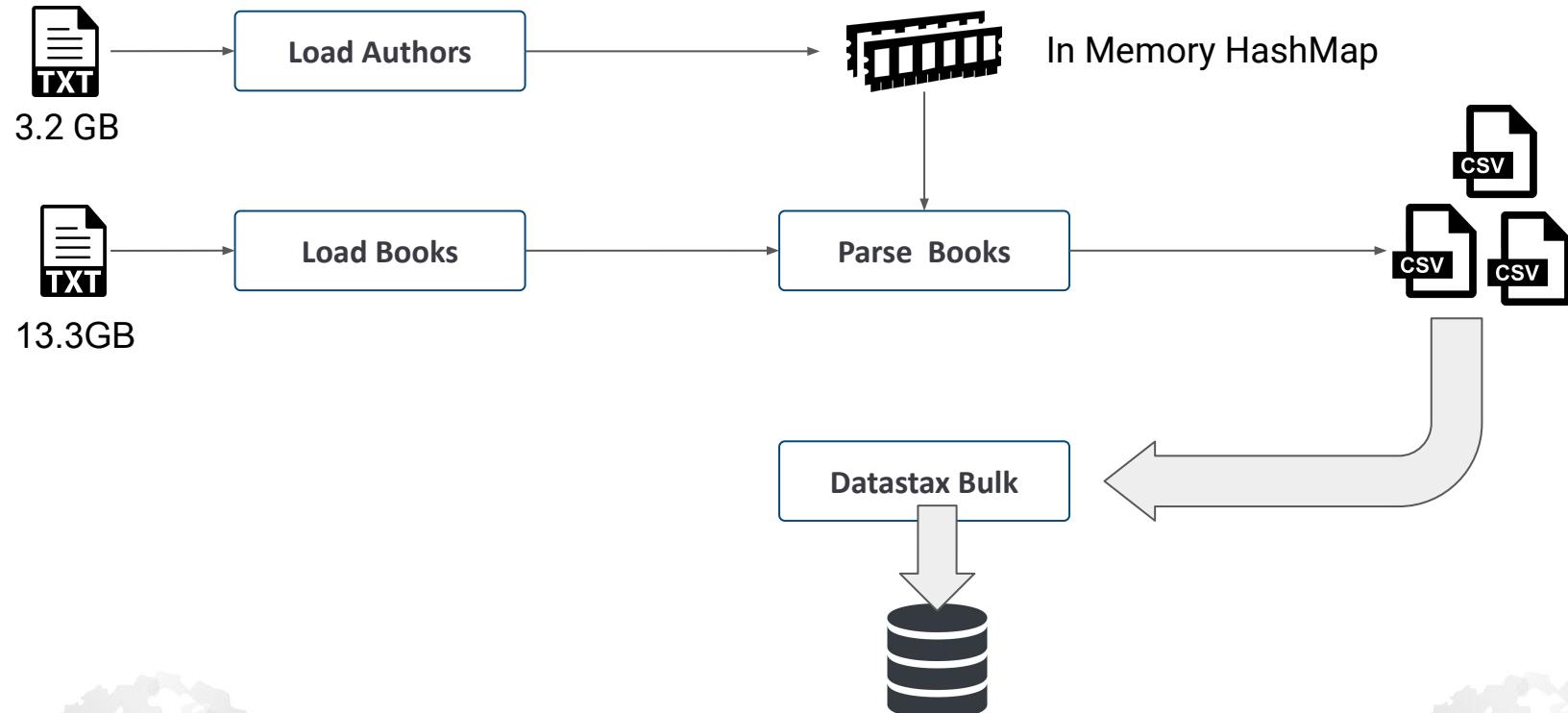


```
id,author_id,author_names,book_description,book_name,cover_ids,published_date
OL10012695W,"[\"OL3979418A\", \"OL3979418A\"]",[],Les écrits Bick,"[\\"3162295\\"]",2009-12-11
OL1000768W,"[\"OL92956A\"]",[],Spooksville - The Wicked,"[\"4648894\", \"4973718\", \"9206708\"]",2009-12-09
OL10005850W,"[\"OL3971678A\"]",[],L'espace rural au moyen age,"[\"5418328\"]",2009-12-11
OL10003205W,"[\"OL3968648A\"]",[],Manuel de pilotage de montgolfières,"[\"3146614\"]",2009-12-11
OL10002825W,"[\"OL4954686A\"]",[],Fragmentarium,"[\"3146154\"]",2009-12-11
OL10009701W,"[\"OL3976143A\"]",[],Traitement de données avec Excel,"[\"3158138\"]",2009-12-11
```



Dsbulk unload





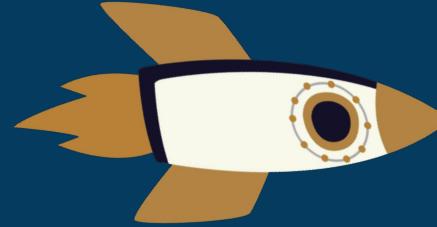
Generate Datastax Bulk Files

```
dsbulk load \
  -c csv \
  -k better_reads \
  -t book_by_id \
  -u token \
  -p ${ASTRA_DB_ADMIN_TOKEN} \
  -b /home/gitpod/.astra/scb_${ASTRA_DB_ID}_${ASTRA_DB_REGION}.zip \
  -maxErrors -1 \
  -logDir /workspace/workshop-betterreads/logs \
  -url /workspace/workshop-betterreads/dataset/book_by_id_0.csv
```



DS Bulk Load





Hands-on (!github)

Working with DATA

- ✓ Start and setup gitpod
- ✓ Work with cqlSH
- ✓ Load Data with DS Bulk



01



Housekeeping
Live and Hands-on

02

System Design
Use case and Technologies

03

Database Setup
Astra, Schema, Data

04

Connect & Use Data
SDK, Spring Data

05

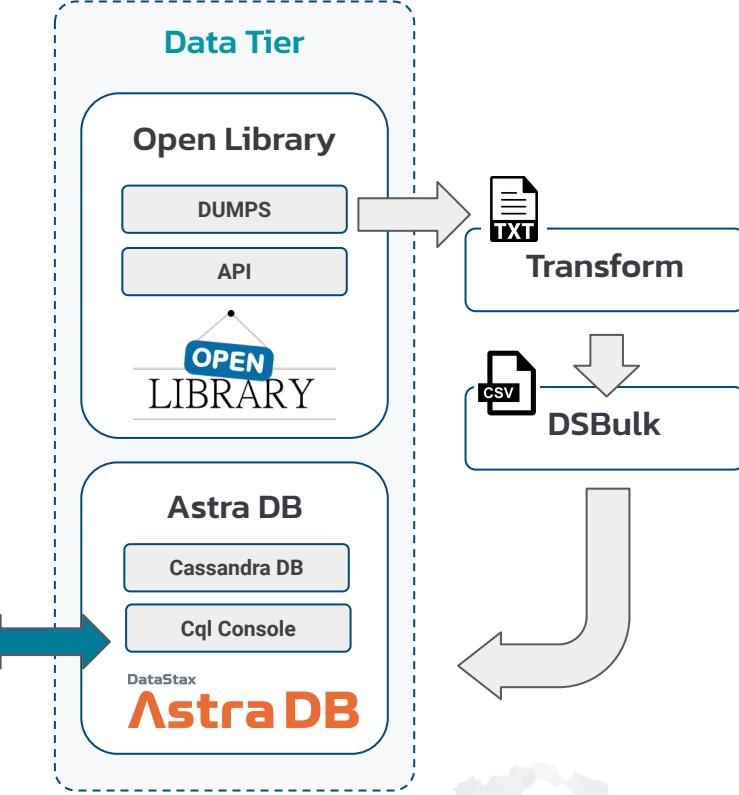
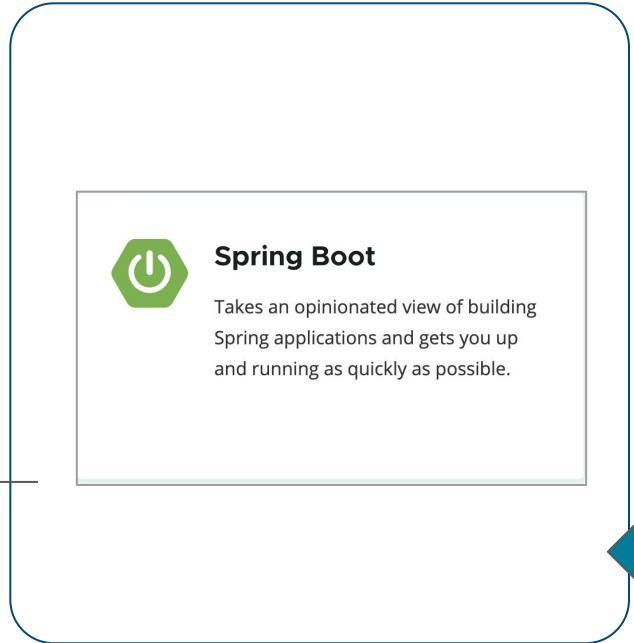
Web and Security
Controller, View, SSO

06

What's next?
Quiz, Homework, Agenda



Agenda



Application Technical Stack

Connectivity

- Token & Datacenter Aware
- Load Balancing Policies
- Retry Policies
- Reconnection Policies
- Connection Pooling
- Health Checks
- Authentication | Authorization
- SSL

Query

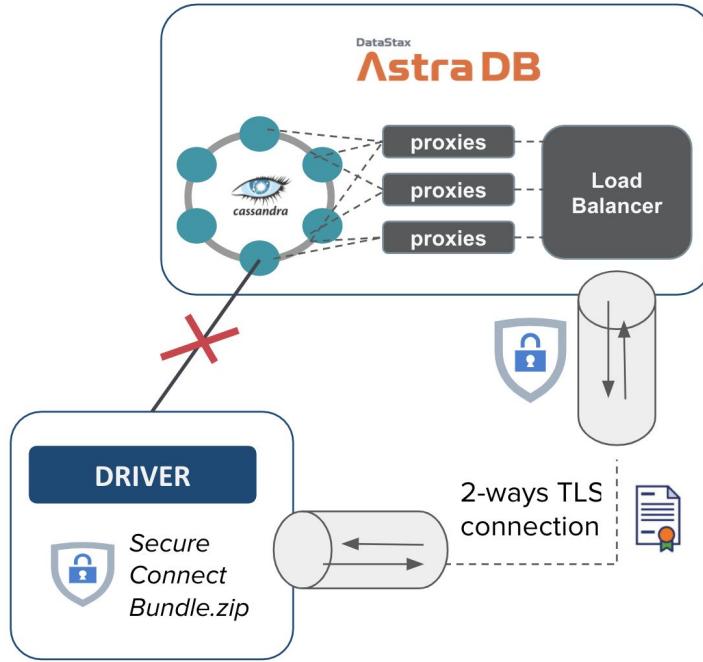
- CQL Support
- Schema Management
- Sync/Async/Reactive API
- Query Builder
- Compression
- Paging

Parsing Results

- Lazy Load
- Object Mapper
- Spring Support
- Paging

CqlSession*: stateful, singleton, pooled

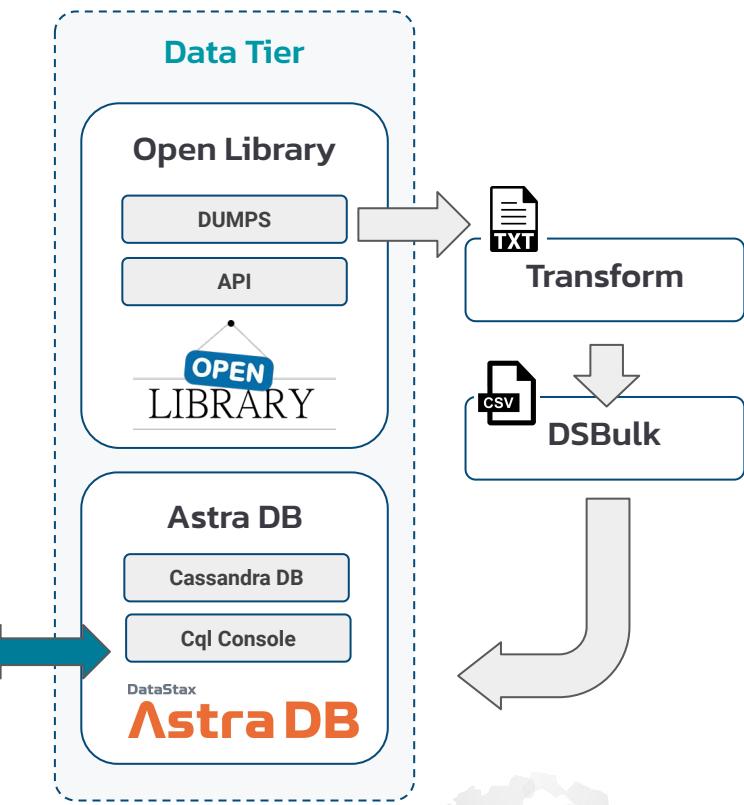
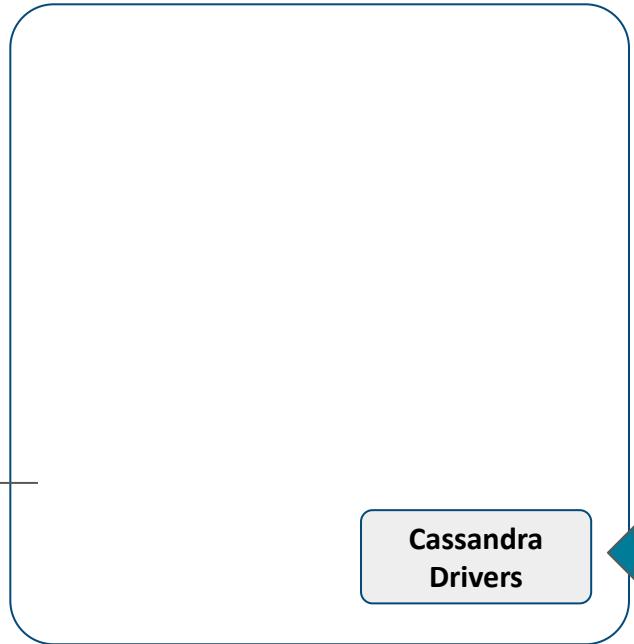




Astra

- Cql Interface
- CqlSession object
- Need SSL configuration
- Need X509 Certificates
- ZIP Secure connect Bundle
- `astra-spring-boot-starter`

Connecting to Astra



Cassandra Drivers and Spring Data

Spring-Data

- CrudRepository<ID, BEAN>
- Working with interface
- Object Mapping
- Working with templates
- Conventions over code
- spring-boot-starter
- autoconfiguration
- keys in application.yaml
- Schema Management

For Cassandra

- Leveraging **CqlSession**
- **CassandraRepository<ID, BEAN>**
- CassandraTemplate, CassandraOperations
- @Table, @PrimaryKey, @Column, @Query

Be “aware”:

- **findAll()**, **@AllowFiltering**,
- Create your own **CqlSession** bean
- Schema creation not recommended
- Data Model first, objects second
- **SimpleCassandraRepository<ID, BEAN>**



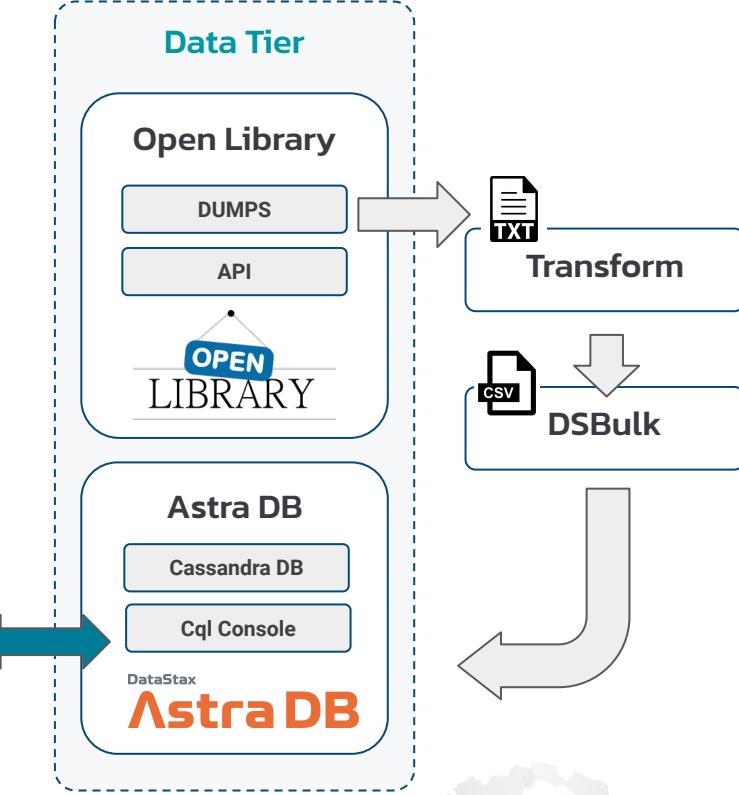
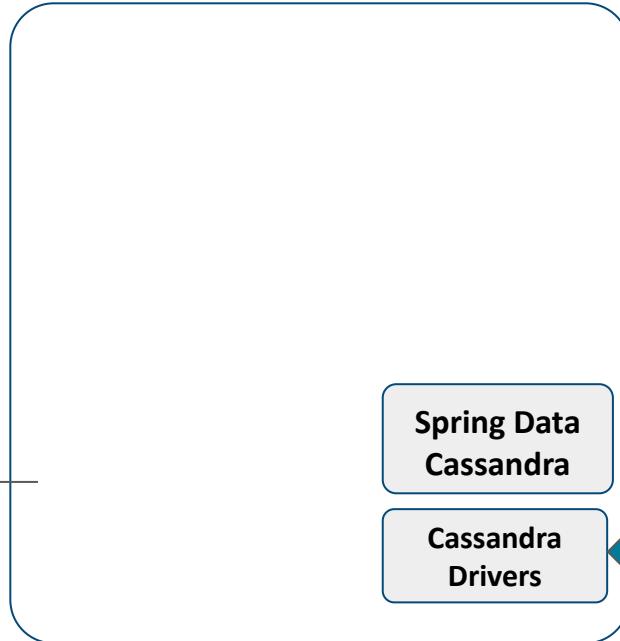
Spring-Data-Cassandra





Spring Data

Provides a consistent approach to data access – relational, non-relational, map-reduce, and beyond.



Spring Data

RestTemplate

- Part of Spring Web
- Java Servlet API
- Blocking Api
- Thread per request
- Context switching = CPU

Spring-WebClient

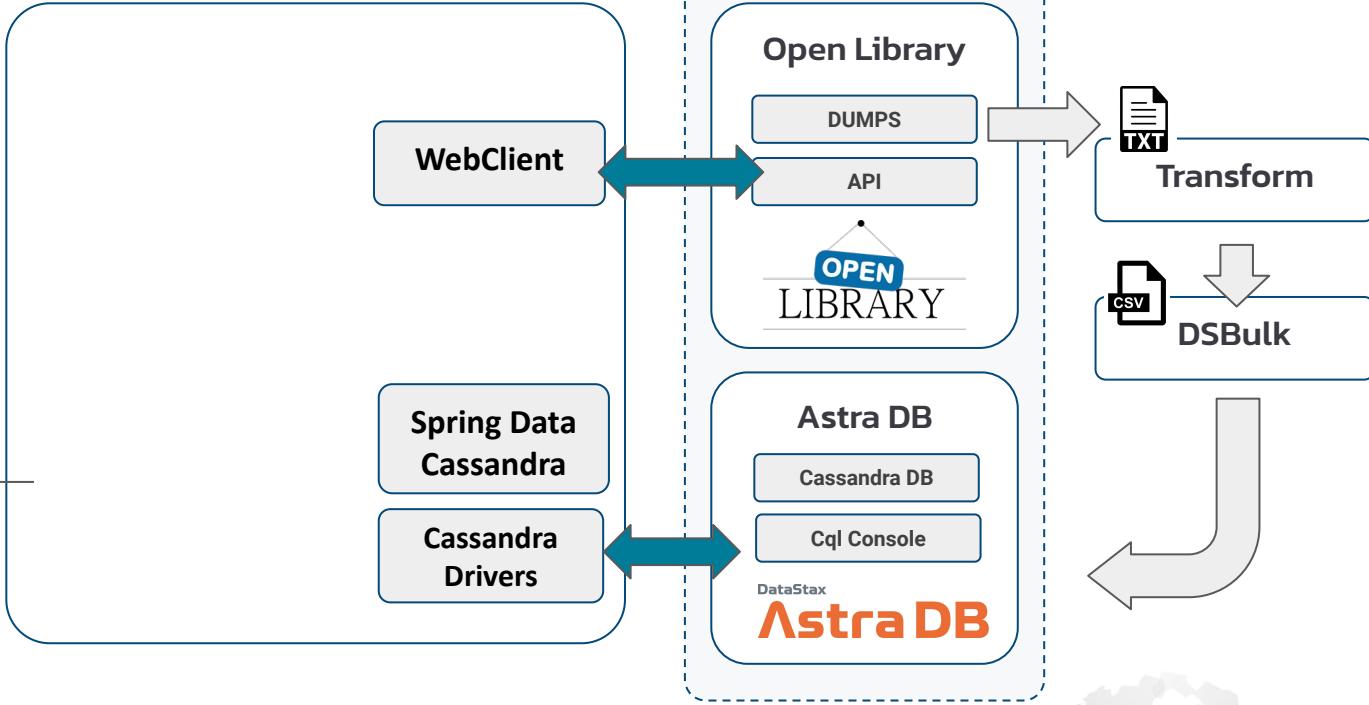
- Part of Webflux (reactive)
- Asynchronous, non blocking
- Build responses with Mono, Flux
- Client Back Pressure



Spring WebClient vs RestTemplate



Spring Web Client



Data Tier

WebClient

Open Library

DUMPS

API

OPEN
LIBRARY

Spring Data
Cassandra

Cassandra
Drivers

Astra DB

Cassandra DB

Cql Console

DataStax
Astra DB



Transform



DSBulk



01



Housekeeping
Live and Hands-on

02

System Design
Use case and Technologies

03

Database Setup
Astra, Schema, Data

04

Connect & Use Data
SDK, Spring Data

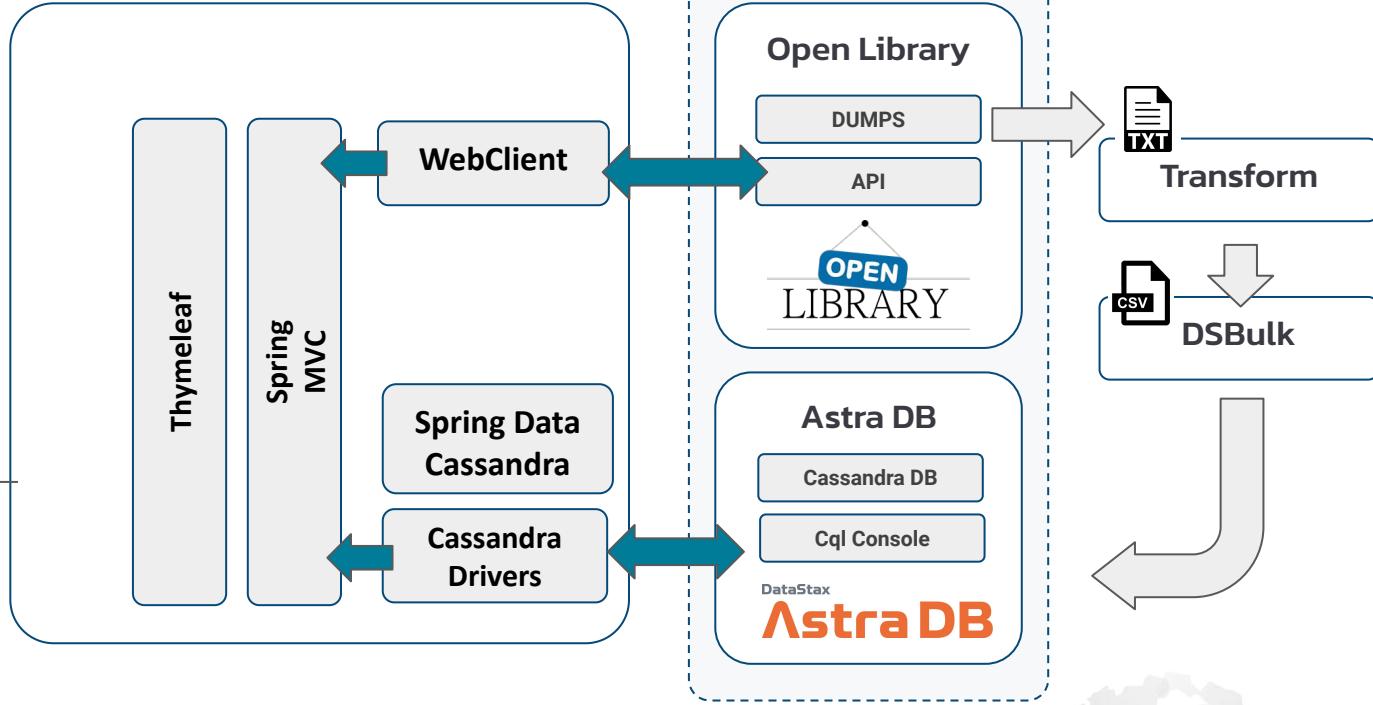
05

Web and Security
Controller, View, SSO

06

What's next?
Quiz, Homework, Agenda



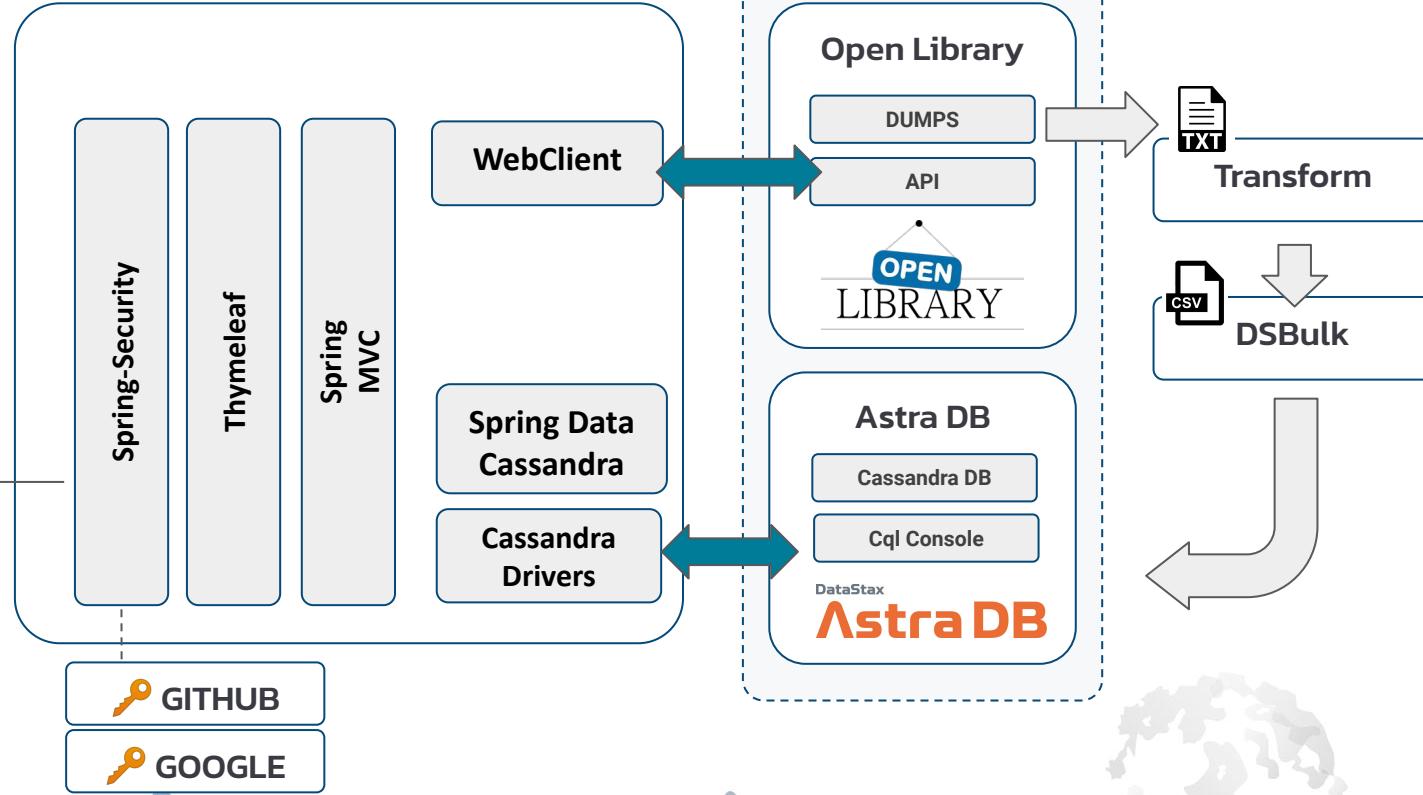


Web Application

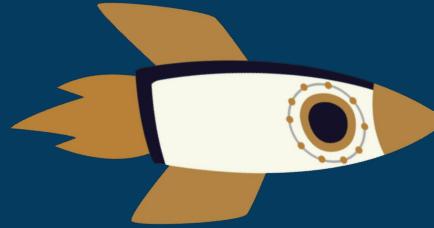


Spring Security

Protects your application with comprehensive and extensible authentication and authorization support.



Authentication



Hands-on (!github)

Web and Security

- ✓ Use application as anonymous
- ✓ Setup Authentication
- ✓ Authenticate and personalize the app



01



Housekeeping
Live and Hands-on

02

System Design
Use case and Technologies

03

Database Setup
Astra, Schema, Data

04

Connect & Use Data
SDK, Spring Data

05

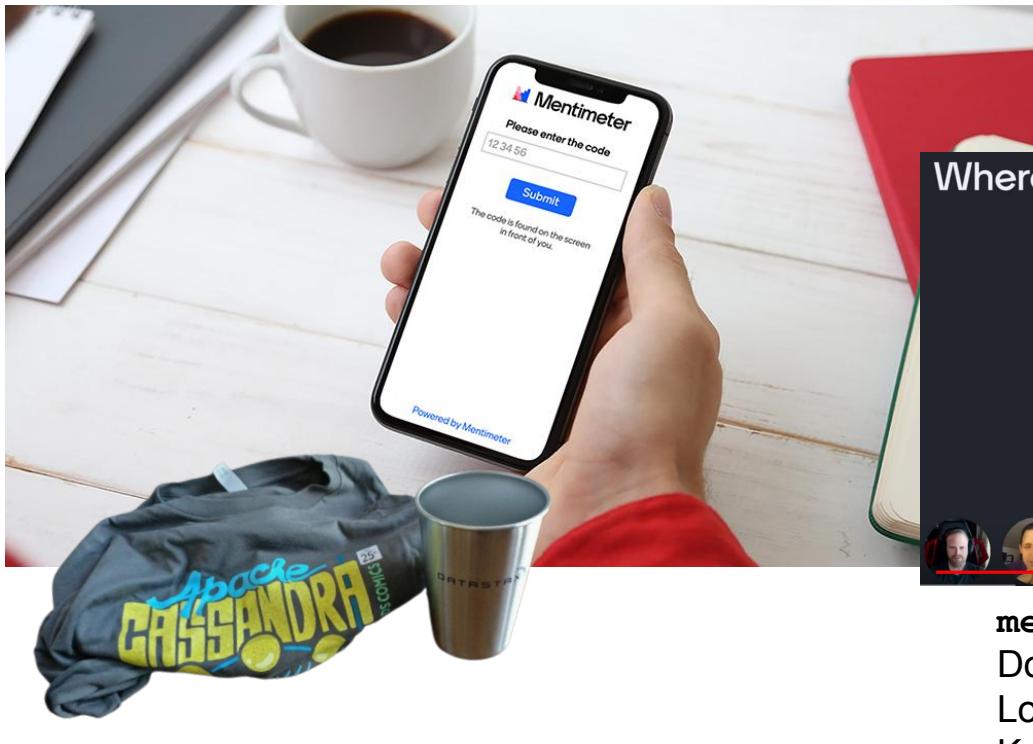
Application MVC
Controller, View, SSO

06

Web and Security
Quiz, Homework, Agenda



Agenda



Where are you from?



menti.com ⇒ enter code
Don't answer in YT chat
Look at phone (not at YT)
Keep it open for later

"Menti" for survey and quiz !

SWAG WINNERS



Congratulations to 1st, 2nd and 3rd place on the Menti quiz!

To claim your prize, please send an email to:

gary.harvey@datastax.com

**** Include a screenshot of your Menti screen**



Swag Winners!

ASTRA DB'S BUILD-A-THON

MAKING SIDE-HUSTLES A REALITY

21 February - 28 May 2022



ASTRA DB'S BUILD-A-THON

MAKING SIDE-HUSTLES A REALITY

21 February - 28 May 2022



JOIN OUR ASTRA DB BUILD-A-THON HACK!

📍 3 months, 3 rounds of challenges. 📍
Join 1 month, 2 months or all 3

Each month, we'll reveal a fresh new set of challenges you can partake in.

All you have to do is have Astra DB as your backend.

USD\$41,000 worth of prizes



REGISTER - buildathonhack.com



DataStax Developers

workshop-chat

<https://www.youtube.com/watch?v=MuwT5xxFVVI> - Subscribe to mailing list: [http...](http://)

Rechercher

PRESENTER — 1
David Jones-Gilardi

HELPER — 7
012345
AaronP
B1nary
Chelsea Navo
Jeremy Hanna
John Sanda
Patrick_McFadin

EN LIGNE — 560
-samu-
6304-42JB
Aahlya
Abdurahim
abhi3pathi
Abhiis.s
Abhineet
Abirsh

Événements
moderator-only
. WELCOME
start-here
code-of-conduct
introductions
upcoming-events
useful-resources
memes
your-ideas
@the-stage

WORKSHOPS
workshop-chat
workshop-feedback
workshop-materials
upcoming-workshops

ASTRA DB
getting-started
astra-apis
astra-development
sample-applications

APACHE CASSANDRA
Cedrick Lun...

RIGGITYREKT Hier à 21:14
I have a 5 node datacenter, 4 nodes are on dse version 5.1.20, one is on dse5.0.15. I am doing some mixed version testing for a class and the one node that is 5.0.15 is coming up as an analytics workload. I dont have /etc/default/dse, instead I am using /etc/init.d/dse-cassandra.
how do I make that node start in cassandra workload, not in analytics?

RIGGITYREKT Hier à 23:39
Okay I found out my issue, when i started DSE 5.0.15 it had endpointsnitch set to DseSimpleSnitch, the rest of my cluster is using PropertyFileSnitch, when i change it to PropertyFileSnitch, it still uses the simple snitch config. looking at the docs i see there is a way to go to GossipingPropertyFileSnitch, but i need the property file one. I can wipe this dbs, do anything with this node to get this done. how do i fix this?
@here

Erick Ramirez Aujourd'hui à 02:19
mixed versions isn't supported and you're guaranteed to run into weird issues that will cause further problems down the track

Cedrick Lunen Aujourd'hui à 09:01
When you start a node you have parameters -k for analytics, -g for graph and -s for search. To remove analytics check and remove -k

Envoyer un message dans #workshop-chat

!discord

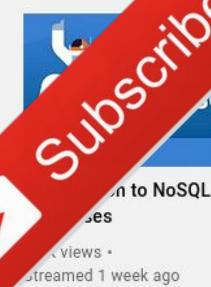
dtsx.io/discord



Datastax Developers Discord (18k+)



Subscribe



Subscribe



Astra Streaming Demo
177 views • 2 weeks ago

Kubernetes Ingress Management with Traefik...
496 views • Streamed 2 weeks ago

Build your own TikTok clone!
1.9K views • Streamed 3 weeks ago

Build your own TikTok Clone!
4K views • Streamed 3 weeks ago

How to use the Connect Driver in Astra DB
113 views • 4 weeks ago

How to use the CQL Console in Astra DB
39 views • 4 weeks ago



How to create an Authentication Token in...
37 views • 4 weeks ago

How to use the Data Loader in Astra DB
62 views • 4 weeks ago

Astra DB Sample App Gallery
36 views • 4 weeks ago

How to use Secure Connect in Astra DB
42 views • 4 weeks ago

Cassandra Day India: CL Room (Workshops)
2.4K views • Streamed 4 weeks ago

Cassandra Day India: RF Room (Talks)
1.3K views • Streamed 1 month ago

Thank You!

