



# Welcome to “Introduction to NoSQL” ~~Webinar~~ Workshop !





# Welcome to our Data Tribe!



# Who is She Loves Data?



## Mission

We empower women with tech & data education and through the data sisterhood, a global, diverse and inclusive support network.



We envision a world where women co-drive the ongoing tech revolution on equal terms.

## Vision

# REACH AND IMPACT



WOMEN TRAINED  
IN 140+ EVENTS



Followers in the  
ONLINE COMMUNITY  
(Cumulative, not uniques.)



MEMBERS /  
SUBSCRIBERS



CHAPTERS  
in 16 countries  
More coming up!

# Our Workshops



## Intro To Data Analytics

- › Intro to data world
- › Introduction to SQL, with hands on session
- › Data visualization & Business Insights, intro session of one of the BI Tools (Tableau, Yellowfin, Power BI..) with hands on session



## Meetups

- › Data is The New Black
- › The Future Data Analytics in Marketing
- › Let's talk about ethics in data.
- › Forget about data analytics let's talk about storytelling.
- › Data engineering



## Soft Skills

- › Presenting with Impact
- › Critical Thinking
- › Communicating with Influence
- › Adopting a growth mindset
- › Build your personal story
- › Public speaking



## Tech courses

- › Introduction to Python
- › R & SQL programming
- › Advanced Data visualization

## Martech for Data-driven marketers

Overview of the digital marketing landscape from a data & measurement perspective



# 2020 the state of the union



she loves data

### Total events

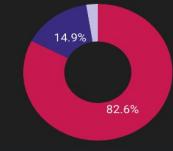
42

Total Registrations  
10,713

Registrations Top Countries	Registered	%
1. Singapore	2,936	27.41%
2. Australia	1,774	16.56%
3. Philippines	1,429	13.34%
4. Indonesia	953	8.9%
5. India	676	6.31%
6. Malaysia	395	3.69%
7. United States of America	383	3.58%
8. Hong Kong	277	2.59%
9. Czechia	270	2.52%
10. Viet Nam	221	2.06%
11. Armenia	201	1.88%
12. New Zealand	188	1.75%
13. South Africa	132	1.23%
14. United Kingdom	111	1.04%
15. United Arab Emirates	87	0.81%
16. Thailand	55	0.51%
17. China	41	0.38%
18. Taiwan	38	0.35%
19. Slovakia	35	0.33%
20. Canada	26	0.24%

1 - 20 / 110 < >

### Attendees background



Female 82.6%  
Male 14.9%  
Do not want to disclose 2.5%

Age Bracket

Age Bracket	%
26 - 30	22%
31 - 35	18%
21 - 25	17%
36 - 40	15%
41 - 45	10%
46 - 50	6%
15 - 20	5%
51 & above	3%

### TOP 5 Industries the people work in

Industry	%
Computer (Technology & Software)	24.05%
Other	21.81%
Financial services (incl. Insurance)	9.97%
Education	9.77%
Information	4.2%

### What is the functional role of your latest jo...

Role	%
Other	17%
Marketing/PR	11.45%
Administrative/Management	9.94%
Data Analytics	9.18%
Research and Development	6.17%
Currently I'm not working (any reason)	5.53%

\*Created in Google Data Studio

6

# Our community in detail



## Professional background

Job Level	%	Job Function	%
Intermediate: Consider themselves as a professional and can work independently.	33,91%	Marketing/PR	11,45%
Mid level management: Have an opportunity to lead teams and divisions.	16,19%	Data Analytics, Data Science and Engineering	9,18%
Entry level: Starting point of (new?) career!	15,72%	Currently I'm not working (any reason)	5,53%
Student (MBA/undergrad/postgrad).	11,07%	Business Development and Sales	3,51%
First level Management: Have the opportunity to lead and coach up and coming individuals as a head specialist.	6,43%	Administrative/Management	9,94%
On a career break (out of work, maternity leave, secondment).	4,72%	Research and Development	6,17%
Self employed: Run your own show and take orders from no one!	4,21%	Finance and Accounting	3,44%
		Operations	4,5%
		Software Development	5,39%
		Other	17%

# Who is She Loves Data?



We are a non-profit, volunteering run organisation founded in 2016 in Singapore

We focus on the importance of DIVERSITY, DATA & DIGITAL LITERACY by conducting **free** workshops, seminars, and meetups for women



We help women to embrace data and tech in order to get them ready for the future of jobs and create their local & global support network, the #datatribe

Since 2016 we have trained over 8000 women

# Who is She Loves Data?



## Vision

To inspire more women to pursue careers in Data & Tech and build a community where like-minded women can come together to learn, connect and have fun!



Significantly expand our reach within Asia Pacific and other continents over the next 3 years - in a sustainable way.

## Mission

# Mary Grygleski - Senior Developer Advocate



- Streaming
- Distributed Systems
- Reactive Systems
- IoT/MQTT



Open  
Liberty



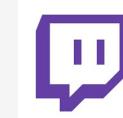
mrgrygles



mary-grygleski



mrgrygles



mrgrygles



she loves data

# Cedrick Lunven - Director of Developer Advocacy



{ REST }

- Trainer
- Public Speaker
- Developers Support
- Developer Applications
- Developer Tooling
  
- Creator of ff4j (ff4j.org)
- Maintainer for 8 years+
  
- Happy developer for 14 years
- Implementing APIs for 8 years



clunven



clunven



clun



she loves data

# Agenda



17:00 – 17:05 Welcome & Intro to the topic

17:05 – 17.10 Housekeeping (Course, Hands-on, Game)

17:10 – 17.25 Introduction: From SQL to NoSQL

17:25 – 17.45 NoSQL Tabular Databases (use cases + exercises)

17:45 – 18.05 NoSQL Document Databases (use cases + exercises)

18:05 – 18.25 NoSQL Key-Value Databases (use cases + exercises)

18:25 – 18.45 NoSQL Graph Databases (use cases + demo)

18:45 – 18.55 Live Quizz

18:55 - 19:00 Final Q&A

Nothing to install !

### Source code + exercises + slides

A screenshot of a GitHub repository page for "DataStax-Examples/todo-astra-javascript-nodejs". The repository has 2 stars, 17 forks, and 28 issues. It contains branches like master, develop, and tags. The commit history shows several commits from "spike.moroz" over the past 10 months, including updates for IAM changes, releases, and bug fixes. A link to "www.datastax.com/register" is also present.

### DataStax Studio



### Database + CQL + APIs



DataStax  
**Astra**



HouseKeeping (Hands-ON)



## Intro to NoSQL Homework

cedrick.lunven@datastax.com [Switch account](#)

Draft restored

The name and photo associated with your Google account will be recorded when you upload files and submit this form. Only the email you enter is part of your response.

\* Required

Email \*

cedrick.lunven@datastax.com

Full Name \*

Your full name (to be displayed on the badge)

Cedrick Lunven



Assessments !



Part #1

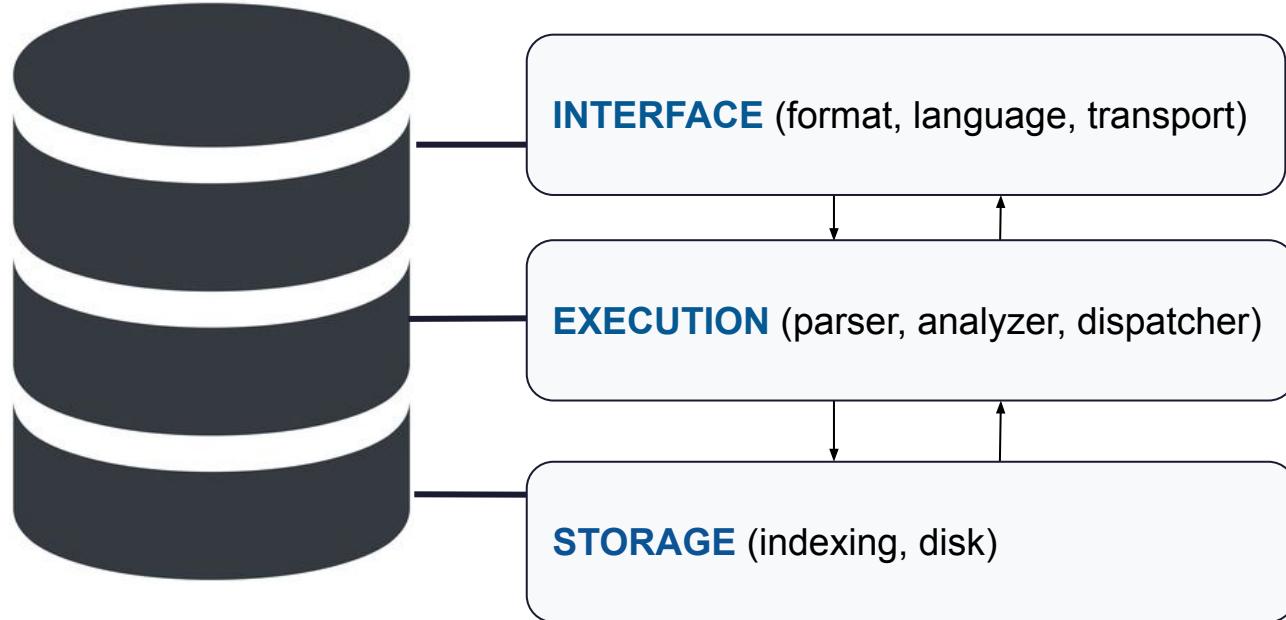
# Introduction to NoSQL Databases



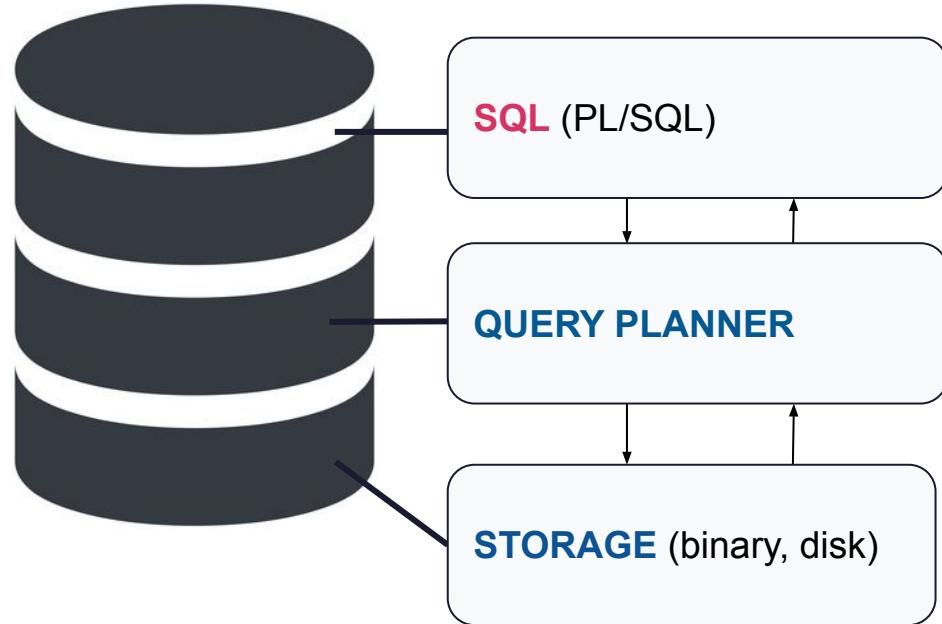
- **Database Management System (DBMS)**
- **Software to save your data (stateful applications)**
- **Software to retrieve your data through queries**
- **Organized collection related data**



What is a Database ?



It is not black box



Relational Databases (**RDBMS**)

## OLTP: Online Transaction Processing

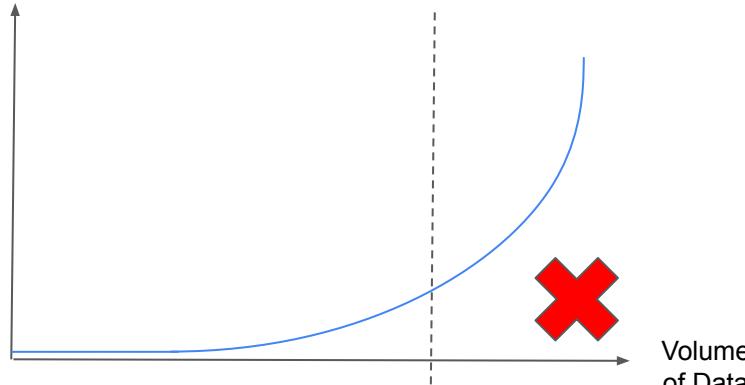
Fast Processing  
Transactional  
High number of transactions  
Hot / Live Data  
“Row oriented”  
Normalized Data (3NF)

## OLAP: Online Analytical Processing

Slow Queries  
Historical  
High volume of data  
Cold Data  
“Column oriented”  
Denormalized Data



Response Time



Relational Databases are **versatile**

- They have been designed to run on a single machine

- New requirements

- V: VOLUME
- V: VELOCITY
- V: VARIETY

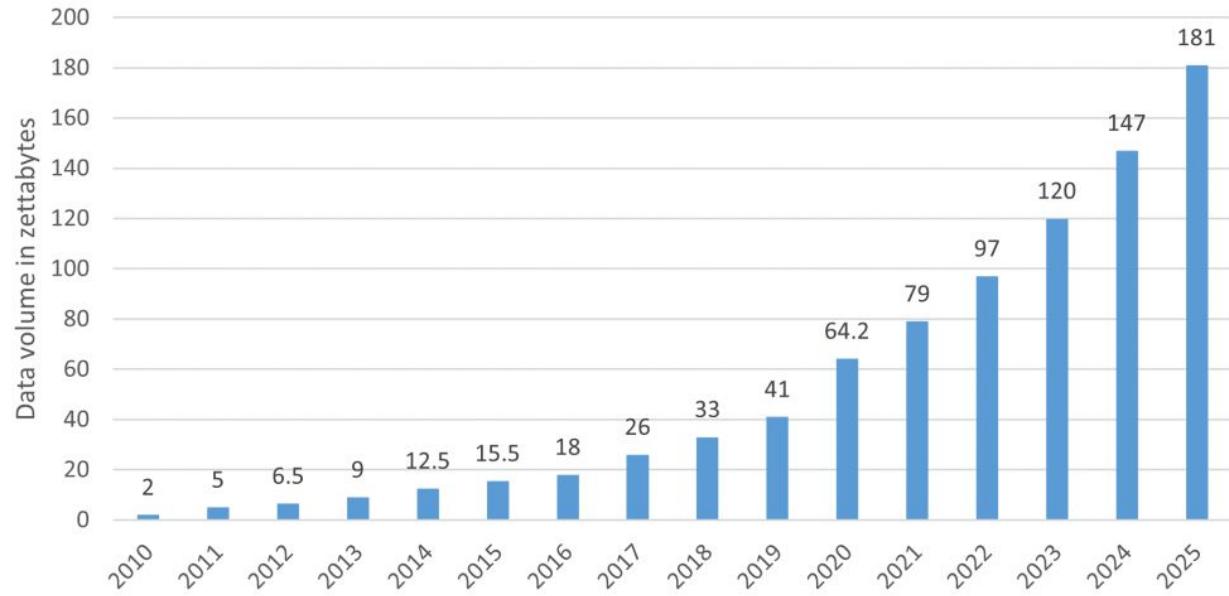
WHAT'S A ZETTABYTE?

1 kilobyte	1,000 000,000,000,000,000,000
1 megabyte	1,000,000 000,000,000,000,000
1 gigabyte	1,000,000,000 000,000,000,000
1 terabyte	1,000,000,000,000 000,000,000
1 petabyte	1,000,000,000,000,000 000,000
1 exabyte	1,000,000,000,000,000,000,000
1 zettabyte	1,000,000,000,000,000,000,000,000

SOURCE: CISCO

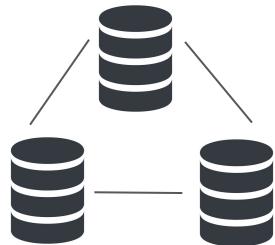


Volume of data created and replicated worldwide (source: IDC)



Relation Databases have limited scalability

- Meetup names on June 11, 2009 in San Francisco
- Web Giants needed more scalable systems
  - Distributed by design (horizontal scalability)
  - Data is replicated
  - Cope with the variety of data



NoSQL

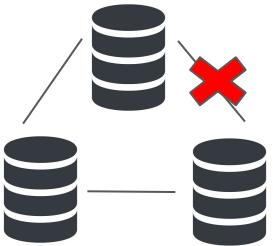


DataStax Developers

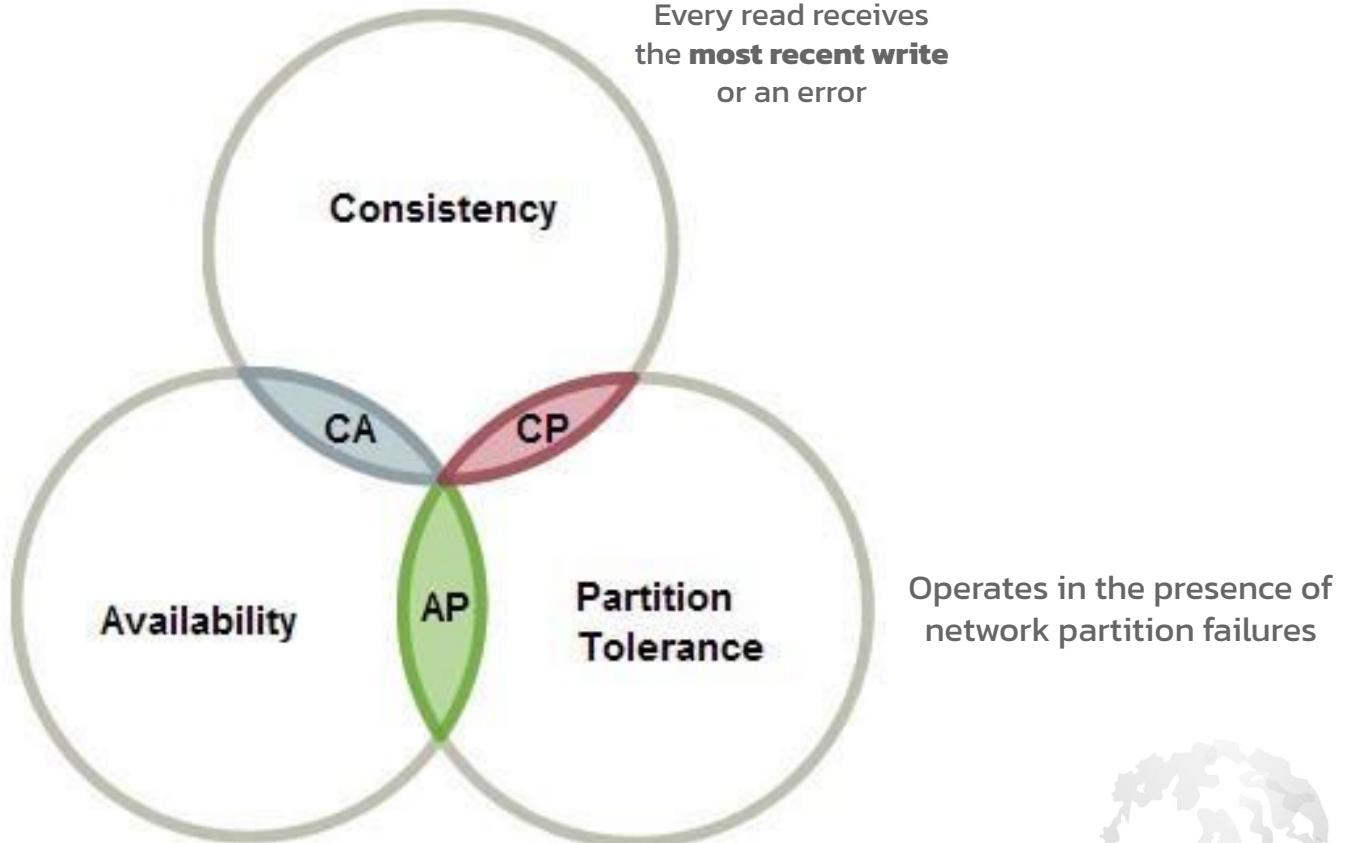


she loves data

Introducing NoSQL



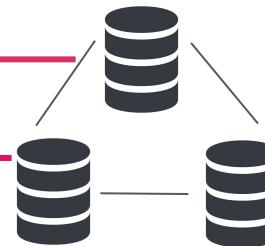
**Always responds,**  
may not always  
return the most recent write



## The CAP Theorem

AP vs CP (CA = not distributed)

Server vs VM vs Container vs Cloud



SQL or JSON or CQL or N1QL or Cypher...

Query Parser and execution engine

Text, Binary, Graph,

Ledger Databases

*awsqldb, ksql\_db*

Time-Series Databases

*Influx, OpenTSDB, Prometheus*

Tabular Databases

*Cassandra, Hbase, Bigtable*

Object Databases

*S3, Minio, Ceph*

Document  
Databases

*MongoDB, Elastic, DocDb*

Graph Databases

*Neo4j, Datastax Graph, Titan*

Key/Value Databases

*Redis, Dynamo, Memcache*



## NoSQL Databases



## Column Oriented Tabular



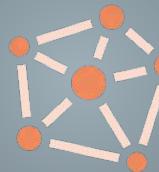
## Document



## Key/value



## Graph



NoSQL Databases Family





# Hands-On

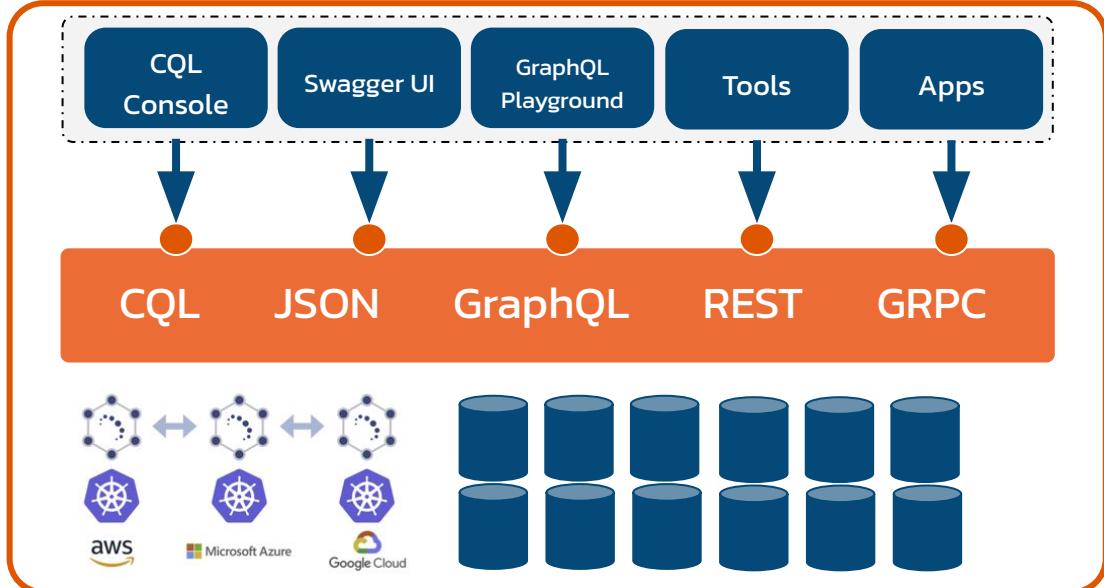
~ AstraDB ~



# Astra DB

**\$25/month credit**

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



**User Interface**  
Web-based developer tools and apps



**STARGATE**

**OSS Stargate.io**  
A data gateway to allow multiple usages



**OSS Apache Cassandra**  
A tabular NoSQL database



Tabular or wide-column database

Astra DB



Document database

Astra DB



Key-value database

Astra DB



Graph database



DataStax Graph



4 Types of NoSQL Databases





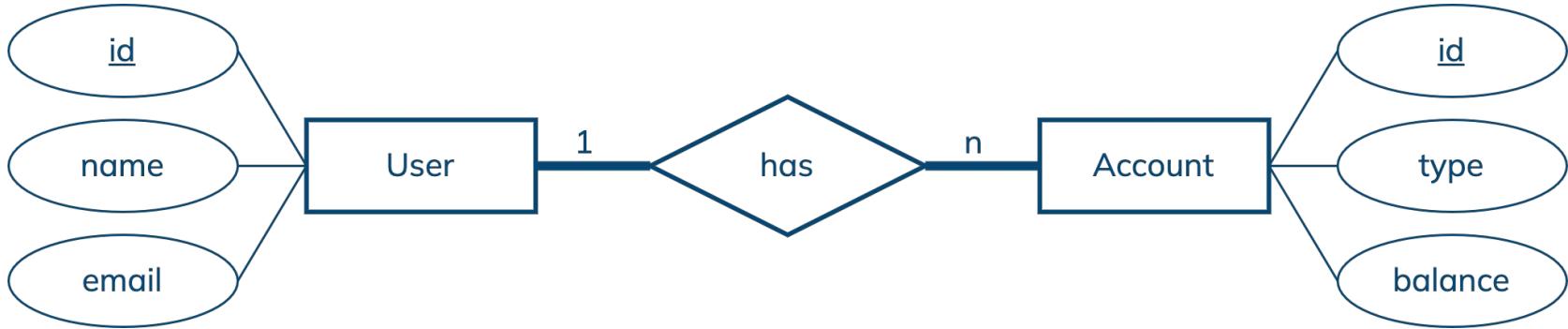
# Lab 1

## Database Setup

[http://github.com/databaxdevs/  
workshop-introduction-to-nosql](http://github.com/databaxdevs/workshop-introduction-to-nosql)

- ✓ Create a new database
- ✓ Wake up an existing hibernated database



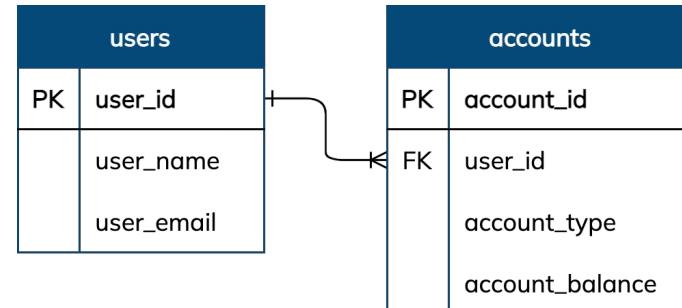


- Meet Alice  [alice@example.org](mailto:alice@example.org) and Bob  [bob@example.org](mailto:bob@example.org)
-  **has** Checking and Savings accounts with balances 2500 and 1500
-  **has one** Checking account with balance 1000

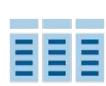
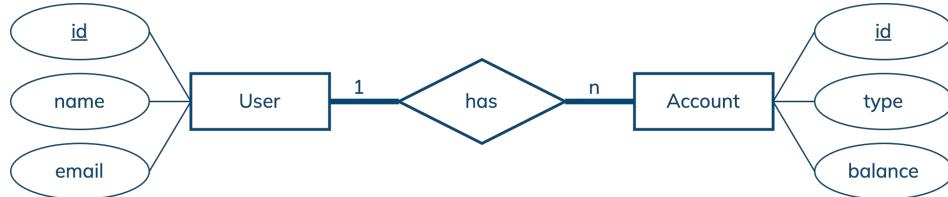
## Running Example: Entity-Relationship Model

users		
user_id	user_name	user_email
1cafb6a4-396c-4da1-8180-83531b6a41e3	Alice	alice@example.org
0d2b2319-9c0b-4ecb-8953-98687f6a99ce	Bob	bob@example.org

accounts			
account_id	account_type	account_balance	user_id
83428a85-5c8f-4398-8019-918d6e1d3a93	Checking	2500	1cafb6a4-396c-4da1-8180-83531b6a41e3
811b56c3-cead-40d9-9a3d-e230dc64f2f	Savings	1500	1cafb6a4-396c-4da1-8180-83531b6a41e3
81def5e2-84f4-4885-a920-1c14d2be3c20	Checking	1000	0d2b2319-9c0b-4ecb-8953-98687f6a99ce



Running Example: Relational Data Model



?



?



?



?



Running Example: What about NoSQL?



## Part #2

# Tabular NoSQL Databases



## Data

- Tables with columns and rows; denormalized tables
- Primary, partition, clustering (sorting) keys
- Secondary indexes, materialized views

## Query

- Retrieving one, some or all rows from a partition
- SQL-like, no joins, indexed columns, lightweight transactions

## Use Cases

- General-purpose; many similarities to relational databases
- Transaction Processing, IoT, Time Series, Messaging, Activity Tracking, Content Management, eCommerce, Retail, Finance



Astra DB



APACHE  
HBASE



```
SELECT account_type, account_balance
FROM accounts_by_users
WHERE user_id = ?;
```

user_id	account_id	user_name	account_balance	account_type
0d2b2319-9c0b-4ecb-8953-98687f6a99ce	81def5e2-84f4-4885-a920-1c14d2be3c20	Bob	1000	Checking
1cafb6a4-396c-4dal-8180-83531b6a41e3	811b56c3-cead-40d9-9a3d-e230dc64f2f	Alice	1500	Savings
1cafb6a4-396c-4dal-8180-83531b6a41e3	83428a85-5c8f-4398-8019-918d6e1d3a93	Alice	2500	Checking



Tabular Databases

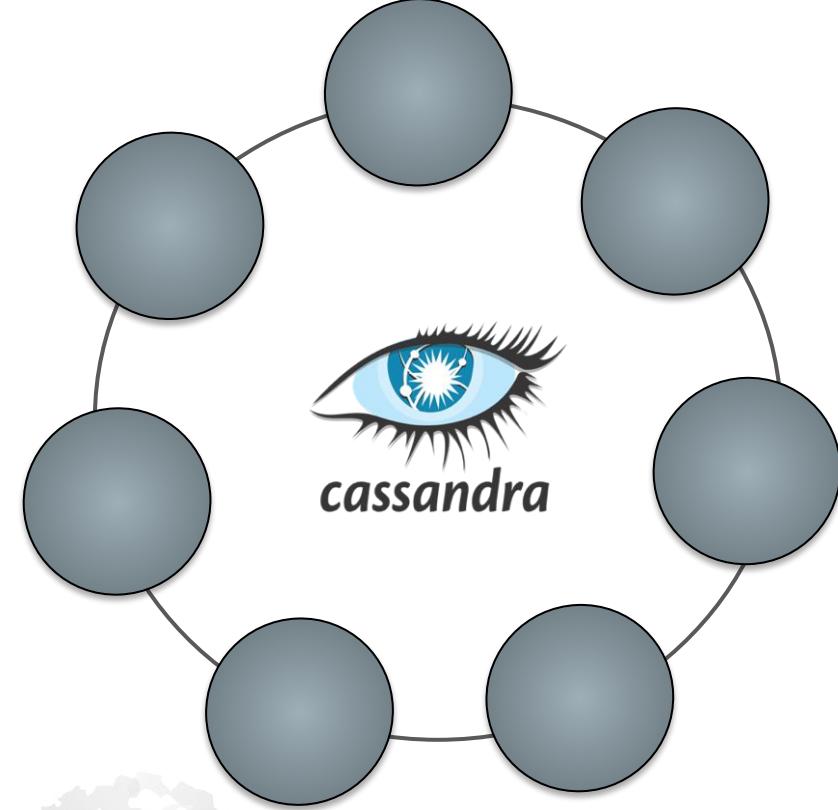


accounts_by_user	
user_id	K
account_id	C $\uparrow$
account_type	S
account_balance	S
use_name	S
use_email	S

Icon of a person with a lightsaber.

accounts_by_user						
	user_id	account_id	account_type	account_balance	user_name	user_email
1cafb6a4-396c-4da1-8180-83531b6a41e3		811b56c3-cead-40d9-9a3d-e230dc64f2f	Savings	1500	Alice	alice@example.org
		83428a85-5c8f-4398-8019-918d6e1d3a93	Checking	2500		
0d2b2319-9c0b-4ecb-8953-98687f6a99ce		81def5e2-84f4-4885-a920-1c14d2be3c20	Checking	1000	Bob	bob@example.org

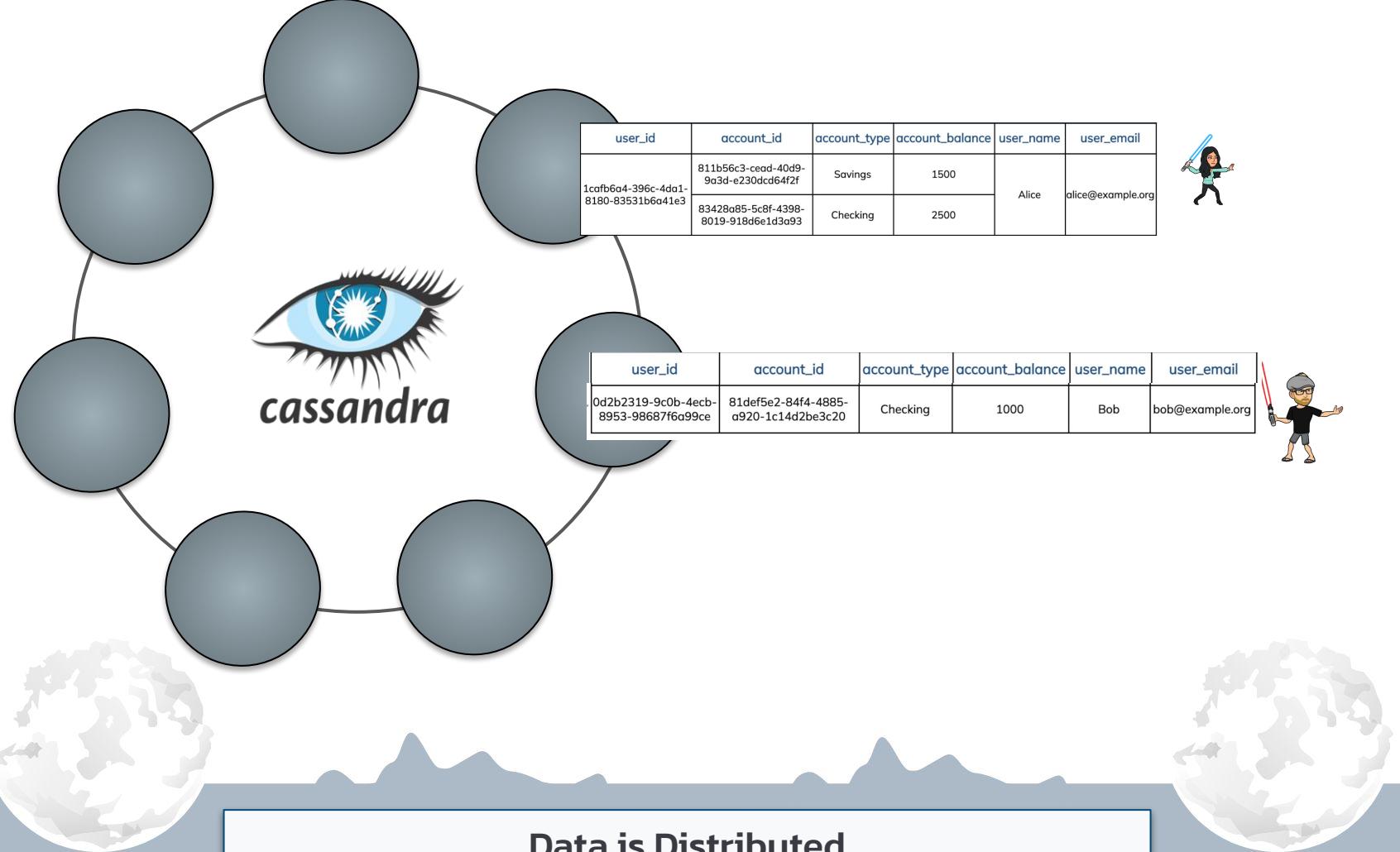
Tabular / Wide-Column Data Model



accounts_by_user					
user_id	account_id	account_type	account_balance	user_name	user_email
1cafb6a4-396c-4da1-8180-83531b6a41e3	811b56c3-cead-40d9-9a3d-e230cd64f2f	Savings	1500	Alice	alice@example.org
	83428a85-5c8f-4398-8019-918d6e1d3a93	Checking	2500		
0d2b2319-9c0b-4ecb-8953-98687f6a99ce	81def5e2-84f4-4885-a920-1c14d2be3c20	Checking	1000	Bob	bob@example.org

Partition Key

Data is Distributed





# Lab 2

## Tabular Databases

[http://github.com/databaxdevs/  
workshop-introduction-to-nosql](http://github.com/databaxdevs/workshop-introduction-to-nosql)

✓ Create Tables

✓ Working with Data and Partitions



## Part #3

# Document NoSQL Databases



## Data

- JSON, semi-structured documents
- Schemaless is a marketing term
- Schema-on-read, schema embedded into the document

## Query

- Search using document IDs and JSON internal field values

## Use Cases

- Flexible schema
- Excellent mapping between database and app data models
- Content Management, Catalogs, Mobile Apps



## users

{ JSON<sub>A</sub> }

{ JSON<sub>B</sub> }



```
db.users.find( { name: "Alice" } )
```



```
GET namespaces/nosql1/collections/users?  
where={"name": { "$eq": "Alice" } }
```



DataStax Developers



# Running Example: Document Data Model

```
{  
  "id": "0d2b2319-9c0b-4ecb-8953-98687f6a99ce",  
  "name": "Bob",  
  "email": "bob@example.org",  
  "accounts": [  
    {  
      "id": "81def5e2-84f4-4885-a920-1c14d2be3c20",  
      "type": "Checking",  
      "balance": "1000"  
    }  
  ]  
}
```



```
{  
  "id": "1cafb6a4-396c-4da1-8180-83531b6a41e3",  
  "name": "Alice",  
  "email": "alice@example.org",  
  "accounts": [  
    {  
      "id": "83428a85-5c8f-4398-8019-91826e1d3a93",  
      "type": "Checking",  
      "balance": "2500"  
    },  
    {  
      "id": "811b56c3-cead-40d9-9a3d-e230dc64f2f",  
      "type": "Savings",  
      "balance": "1500"  
    }  
  ]  
}
```



40



# Lab 3

## Document Database

[http://github.com/databaxdevs/  
workshop-introduction-to-nosql](http://github.com/databaxdevs/workshop-introduction-to-nosql)

- ✓ Create a Token
- ✓ Use Swagger APIs



## Part #4

# Key-Value NoSQL Databases



## Data

- Key-value pairs
- Values can be blobs, strings, sets, maps, JSON docs, etc

## Query

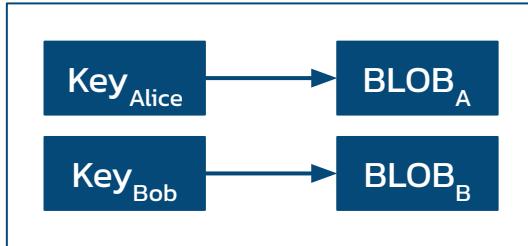
- GET/PUT/DELETE key lookups, limited SQL-like support

## Use Cases

- Best for simple primary key lookups
- Caching, User Sessions, Serving Content by Key



## users



GET  
`/buckets/users_kv/keys/Alice`



SELECT value FROM users\_kv  
WHERE key = 'Alice';  
43



# Running Example: Key-Value Data Model



KV	
key	value
user:1caf6a4-396c-4da1-8180-83531b6a41e3:name	Alice
user:1caf6a4-396c-4da1-8180-83531b6a41e3:email	alice@example.org
user:1caf6a4-396c-4da1-8180-83531b6a41e3:accounts	{ 83428a85-5c8f-4398-8019-918d6e1d3a93, 811b56c3-cead-40d9-9a3d-e230dcd64f2f }
user:0d2b2319-9c0b-4ecb-8953-98687f6a99ce:name	Bob
user:0d2b2319-9c0b-4ecb-8953-98687f6a99ce:email	bob@example.org
user:0d2b2319-9c0b-4ecb-8953-98687f6a99ce:accounts	{ 81def5e2-84f4-4885-a920-1c14d2be3c20 }
account:83428a85-5c8f-4398-8019-918d6e1d3a93:type	Checking
account:83428a85-5c8f-4398-8019-918d6e1d3a93:balance	2500
account:811b56c3-cead-40d9-9a3d-e230dcd64f2f:type	Savings
account:811b56c3-cead-40d9-9a3d-e230dcd64f2f:balance	1500
account:81def5e2-84f4-4885-a920-1c14d2be3c20:type	Checking
account:81def5e2-84f4-4885-a920-1c14d2be3c20:balance	1000

44



# Lab 4

## KeyValue Database

[http://github.com/databricksdevs/  
workshop-introduction-to-nosql](http://github.com/databricksdevs/workshop-introduction-to-nosql)

- ✓ Create kv tables
- ✓ Access KV tables



## Part #5

# Graph NoSQL Databases



# Graph Databases



DataStax Graph



TigerGraph



AllegroGraph

## Data

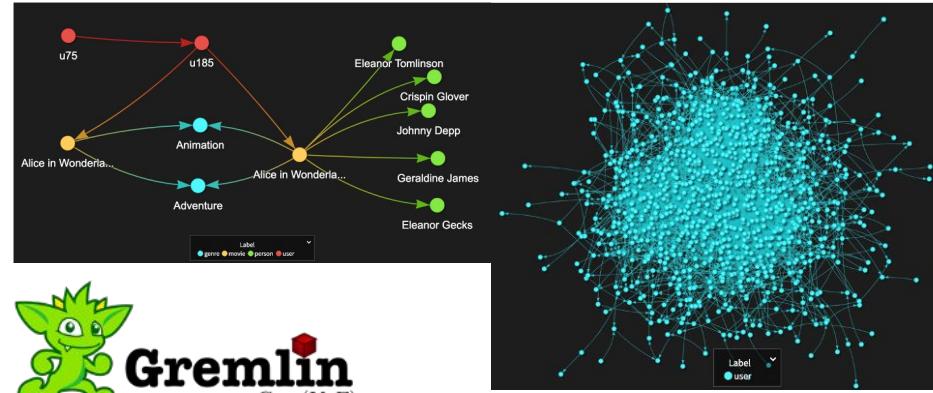
- Vertices, edges, properties
- Property graphs, RDF graphs, knowledge graphs

## Query

- Gremlin, Cypher, GSQL, SPARQL

## Use Cases

- Focus on exploring connections, links and relationships
- Customer 360, Personalization, Fraud detection, Recommendations, Internet of Things, Asset management, Data integration



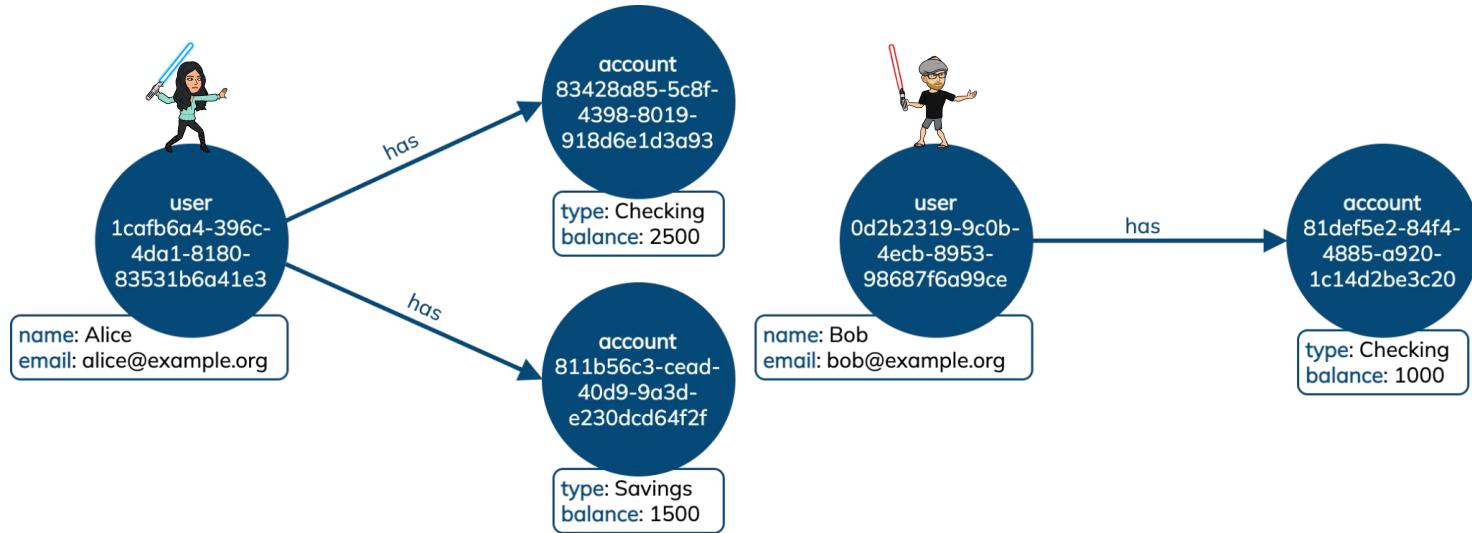
**Gremlin**

$G = (V, E)$

```
g.V().has("user", "name", "Alice").  
repeat(both().simplePath().timeLimit(800)).  
until(has("user", "name", "Bob")).path().limit(1)
```

47



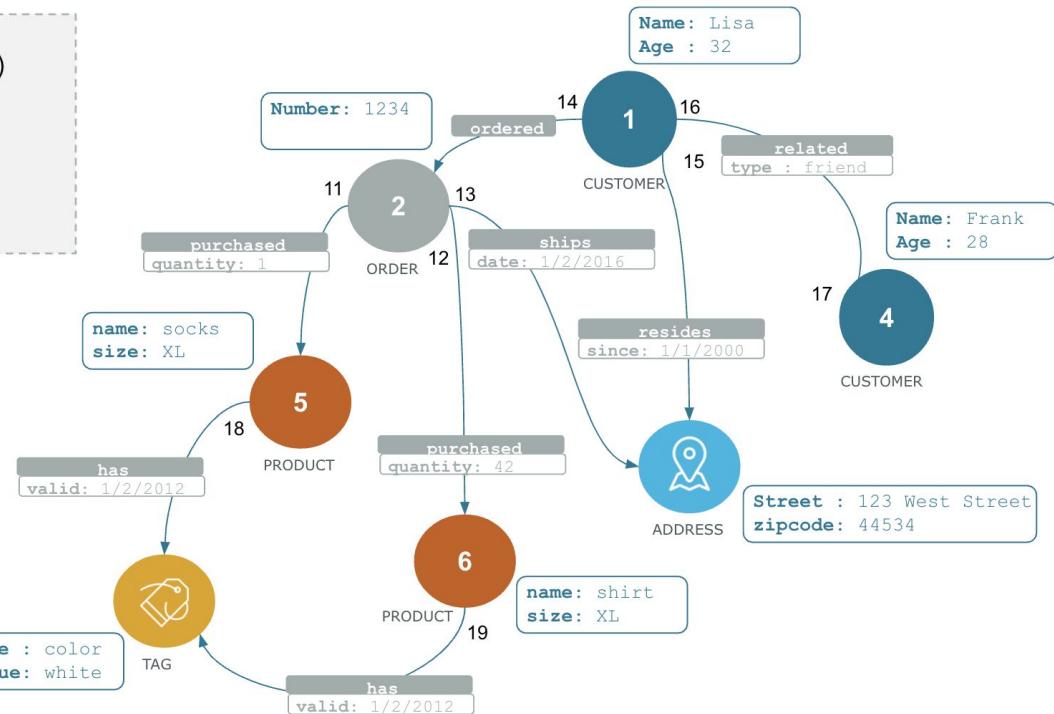


## Running Example: Graph Data Model

```

g.V().has('customer', 'name', 'Lisa')
    .out('ordered')
    .values('number')

```

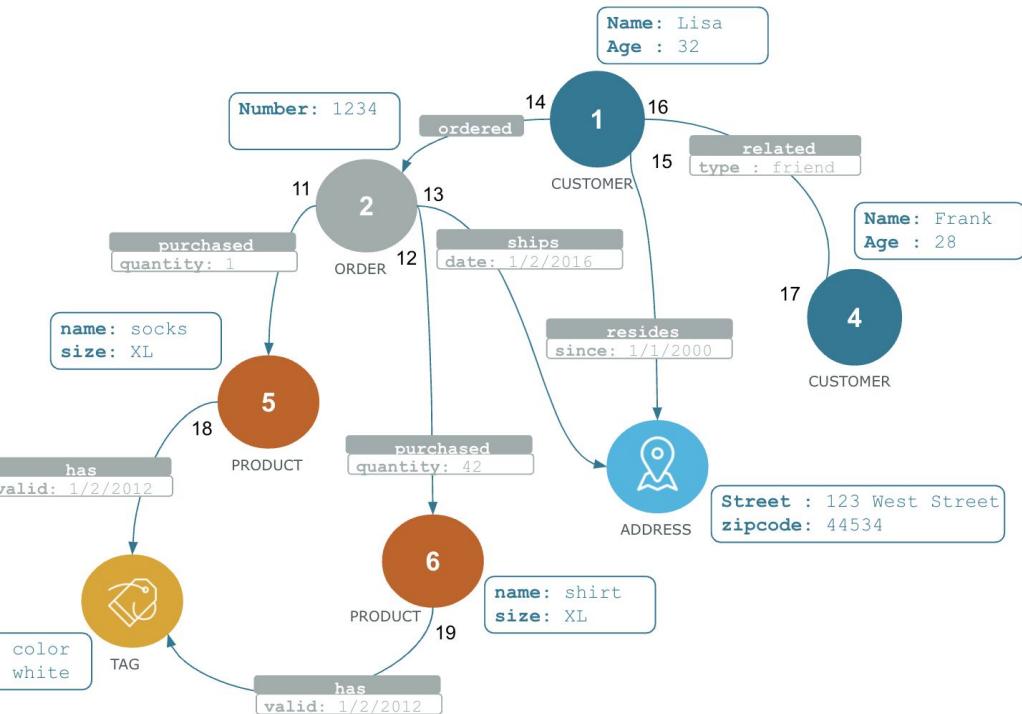


## Find Lisa Friends

```

g.V().has('customer', 'name', 'Frank')
    .outE('related').has('type', 'friend')
    .inV()
    .out('ordered')
    .out('purchased')
    .values('name')

```



Find all product names purchased by Frank's friends.



# Lab 5

## Graph Databases

[http://github.com/databricksdevs/  
workshop-introduction-to-nosql](http://github.com/databricksdevs/workshop-introduction-to-nosql)

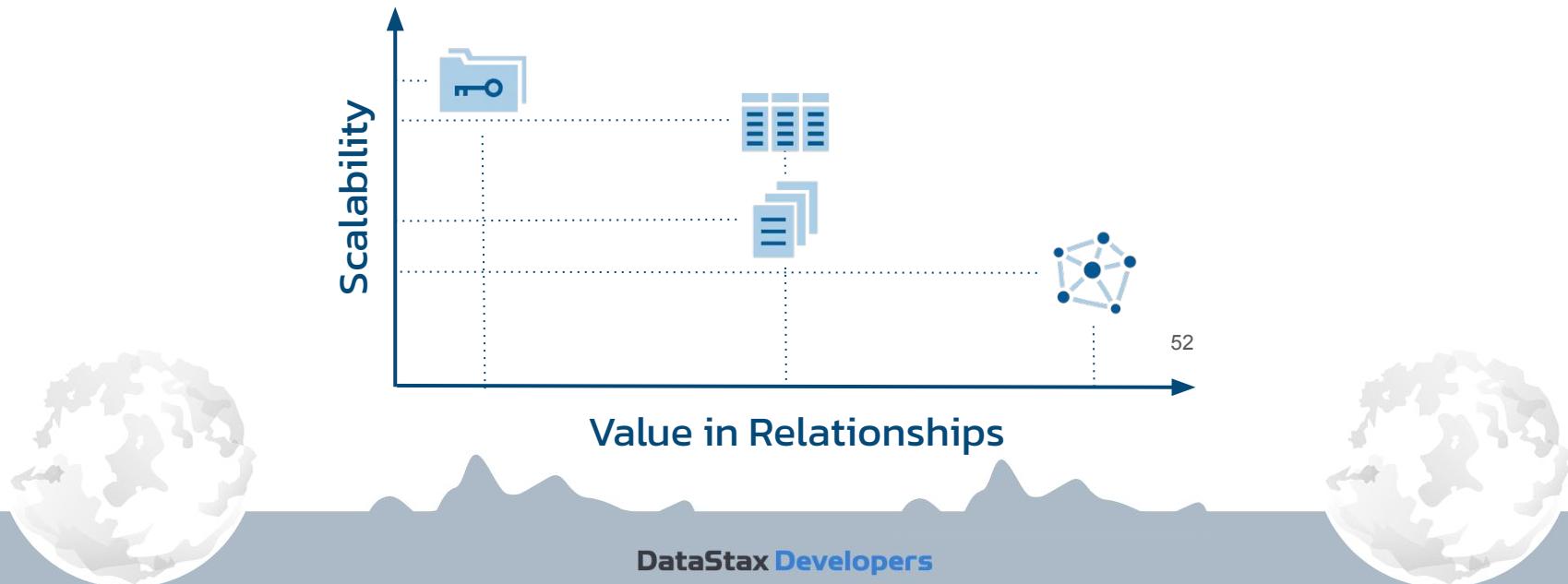
- ✓ Start Studio (docker)
- ✓ Run Sample

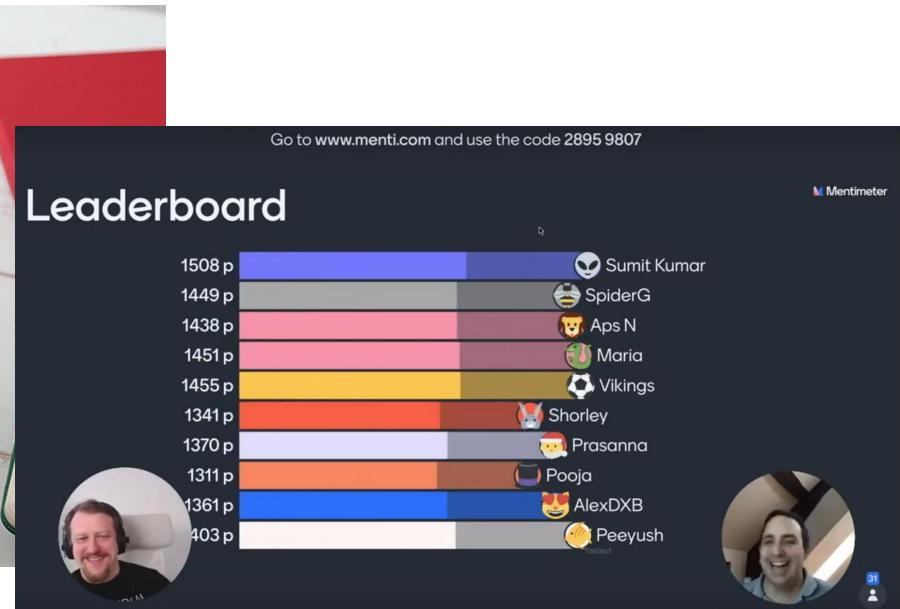
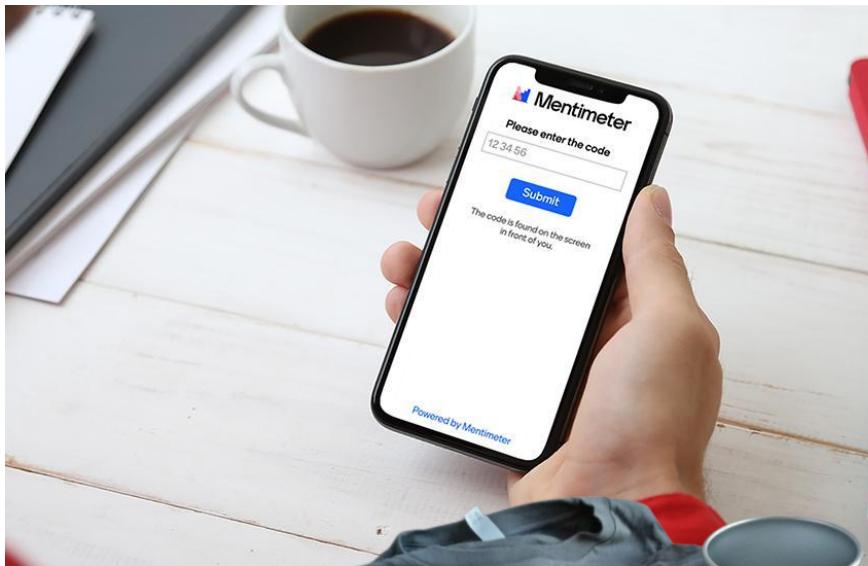


# Should You Choose database?



- Consider a multi-model database, data model is only one aspect
- Use case, performance, scalability, availability, DBaaS, multi-cloud, ...





**menti . com ⇒ enter code  
Don't answer in YT chat  
Look at phone (not at YT)**

**Quiz on Menti.com !**

# 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](mailto:gary.harvey@datastax.com)

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

Swag Winners!



DataStax Developers



## Intro to NoSQL Homework

cedrick.lunven@datastax.com [Switch account](#)

Draft restored

The name and photo associated with your Google account will be recorded when you upload files and submit this form. Only the email you enter is part of your response.

\* Required

Email \*

cedrick.lunven@datastax.com

Full Name \*

Your full name (to be displayed on the badge)

Cedrick Lunven



Assessments !





Join “**She Loves Data**” group &  
follow **@shelovesdata** on LinkedIn



Join “**She Loves Data**” Facebook group  
& follow our **@shelovesdata** page



See some of the snaps  
on the **@shelovesdata** account

**CONNECT WITH  
SHE LOVES DATA**

[www.shelovesdata.com](http://www.shelovesdata.com)