



**CRATE.IO**



INTERNET OF THINGS AND THEIR REQUIREMENTS

THE SOLUTION THAT  
CRATE.IO OFFERS

## PRESENTATION'S OVERVIEW

- The problems that the IoT industry faces
- Features of CrateDB
- Column Store Vs Row Store
- Real use cases of CrateDB
- Common solution architecture the associated complexities and problems
- Similar solution architecture but with the CrateDB this time
- Live demo
- Q&A



# PROBLEMS WITH THE IOT

- Huge amount o data
  - Millions of data per second from a number of different devices, sensors, data centres, logs, cameras, watches, etc
- Data diversity
  - Could be structured (ie. JSON) and unstructured data (ie. images & videos)
- Real time query performance
- Monitoring and alerting
- Complex queries that include terabytes of historic data
- Scalability & growth. All the previous problems are constantly getting bigger

# FEATURES OF CRATEDB (1)

- Scalable. Automatic data rebalancing
- Shared nothing architecture (no special master node)
- Distributed SQL queries aggregations and search
- Highly available
- Real time ingestion
- Any data and BLOBs

# FEATURES OF CRATEDB (2)

- Time series analysis
- Geospatial queries (includes geo\_point and geo\_shape types)

**Dynamic schemas** (no locks needed on tables in order to add new columns or even nested objects)

ie: create table my\_table ( author object( **dynamic** ) as ( name string, birthday timestamp ) );

- Transactional (eventually consistent). Atomic at row level
- Backups
- Openness and flexibility

# COLUMN STORE vs ROW STORE (1)

RowId	EmpId	Lastname	Firstname	Salary
001	10	Smith	Joe	40000
002	12	Jones	Mary	50000
003	11	Johnson	Cathy	44000
004	22	Jones	Bob	55000

- This is a 2D abstraction.
- How the data are written on the disk makes the difference between the two approaches.
- The bottleneck is the I/O. Even a typical SSD is thousands of times slower than RAM.

## Row oriented systems

001:10,Smith,Joe,40000;  
002:12,Jones,Mary,50000;  
003:11,Johnson,Cathy,44000;  
004:22,Jones,Bob,55000;

Row-based systems are not efficient at performing set-wide operations on the whole table, as opposed to a small number of specific records.

## An index on salary col would look like this:

001:40000;  
003:44000;  
002:50000;  
004:55000;

Maintaining indexes adds overhead to the system, especially when new data is written to the database. Records not only need to be stored in the main table, but any attached indexes have to be updated as well.

# COLUMN STORE vs ROW STORE (2)

RowId	EmpId	Lastname	Firstname	Salary
001	10	Smith	Joe	40000
002	12	Jones	Mary	50000
003	11	Johnson	Cathy	44000
004	22	Jones	Bob	55000

<Column Value>:<RowId>,<Column Value>:<RowId>....;

## Column oriented systems

10:001,12:002,11:003,22:004;  
Smith:001,**Jones**:002,Johnson:003,**Jones**:004;  
Joe:001,Mary:002,Cathy:003,Bob:004;  
40000:001,50000:002,44000:003,55000:004;

Optimisation



## Column oriented systems

10:001,12:002,11:003,22:004;  
Smith:001,Jones:**002,004**,Johnson:003;  
Joe:001,Mary:002,Cathy:003,Bob:004;  
40000:001,50000:002,44000:003,55000:004;

Query: "find all the people with the last name Jones", the answer is retrieved in a single operation. Other operations, like counting the number of matching records or performing math over a set of data, can be greatly improved through this organisation.

Operations that retrieve all the data for a given object (the entire row) are slower.

These whole row operations are generally rare



# REAL USE CASES OF CRATEDB

- [gantner-instruments.com](http://gantner-instruments.com): Wind turbines failure prediction
- [clickdrive.io](http://clickdrive.io): 2k data points per car/second
- [skyhighnetworks.com](http://skyhighnetworks.com): monitors logs, clouds, data centres etc
- [roomonitor.com](http://roomonitor.com): Monitors noise, possible incidents, a/c etc



The Clickdrive website banner features a light gray background with a faint map of a city. Several car icons are scattered across the map, with one car in the center having a green circular signal around it. The Clickdrive logo is in the top left corner, and the main headline is centered.

**Clickdrive**  
Evolution of connected cars

## Manage your fleet with Clickdrive

The most advanced solution for Connected Cars.



The Skyhigh website banner has a blue background with a blurred image of a person's hand typing on a laptop. The Skyhigh logo is in the top left, and a navigation menu is in the top right. The main headline is centered.

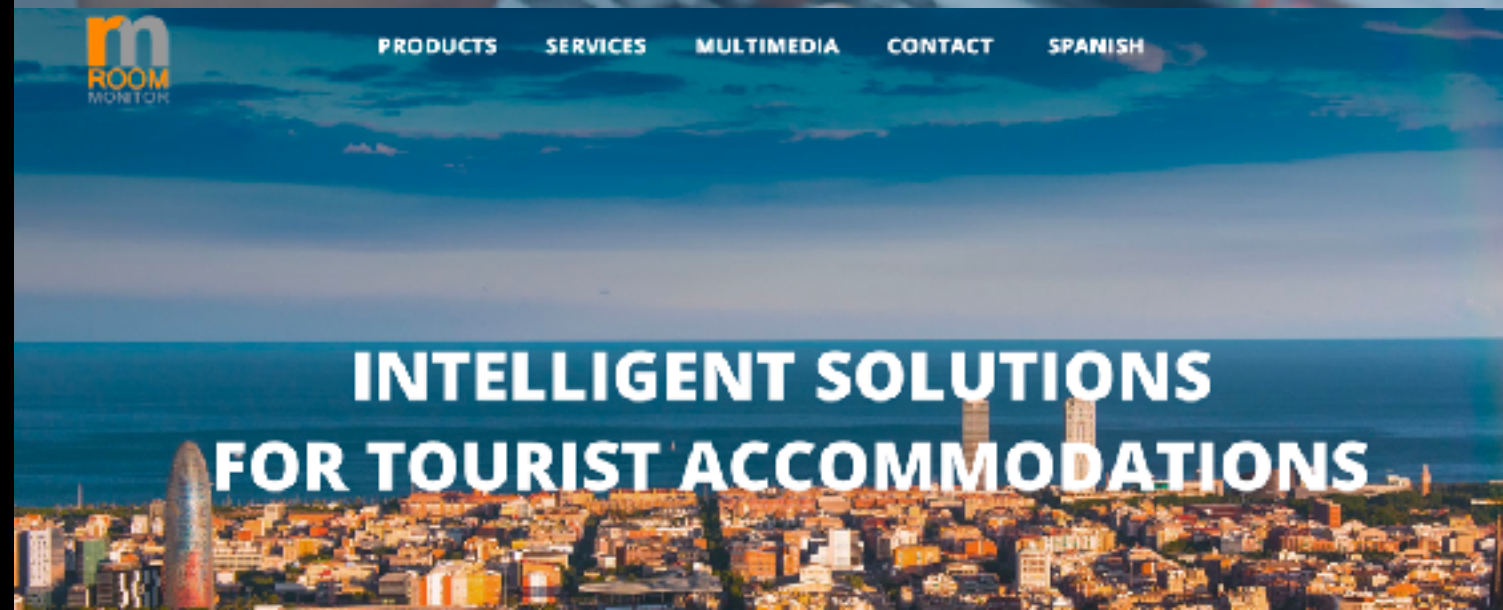
skyhigh

Product ▾ Customers ▾ Partners ▾ Resources ▾ Company ▾

BLOG CONTACT

## Cloud Access Security Broker

Skyhigh enables organizations to enforce security, compliance, and governance policies across all cloud services, all users, and all devices



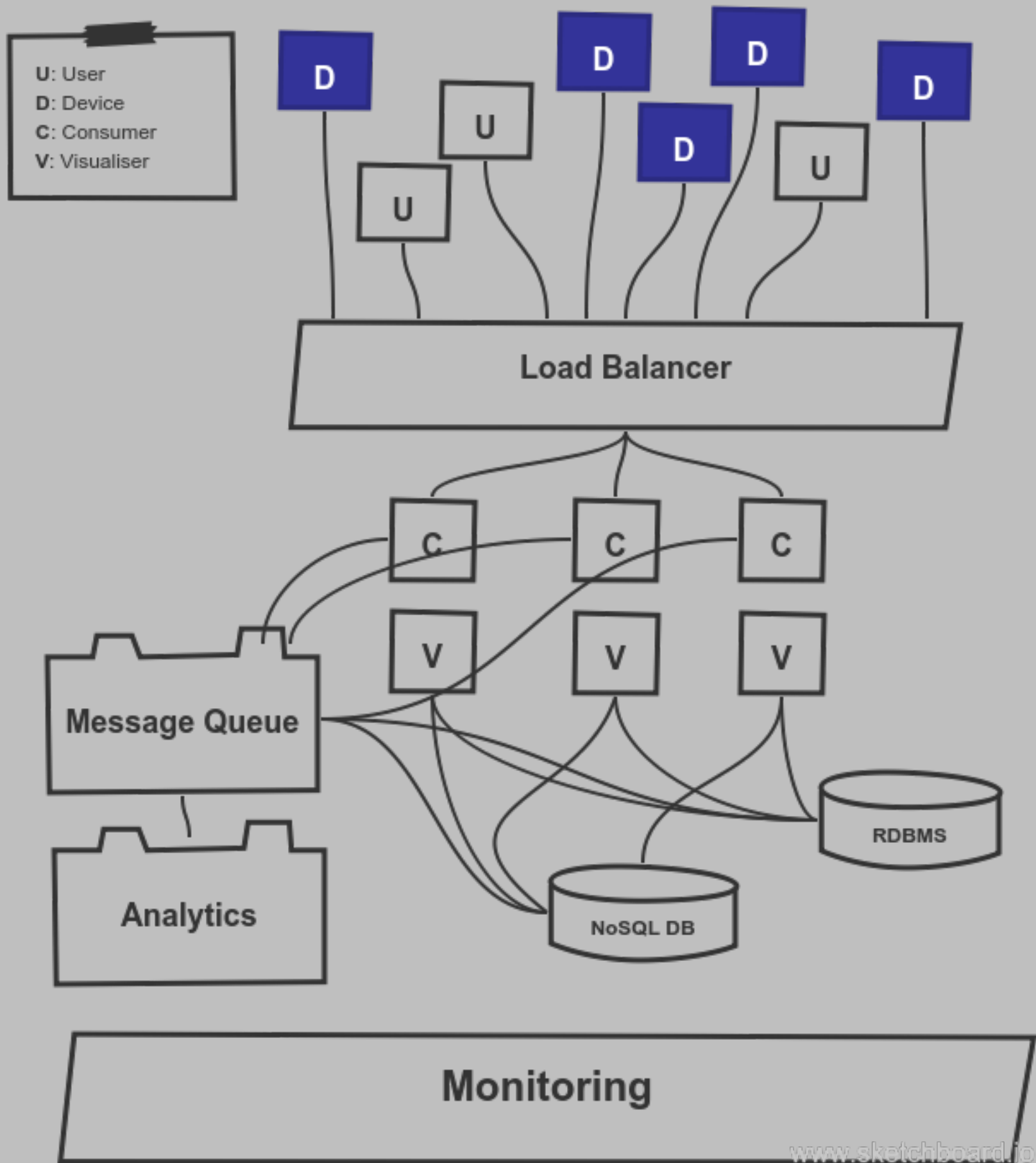
The Roomonitor website banner features a vibrant image of a city skyline at sunset. The Roomonitor logo is in the top left, and a navigation menu is in the top right. The main headline is centered at the bottom.

**m ROOM MONITOR**

PRODUCTS SERVICES MULTIMEDIA CONTACT SPANISH

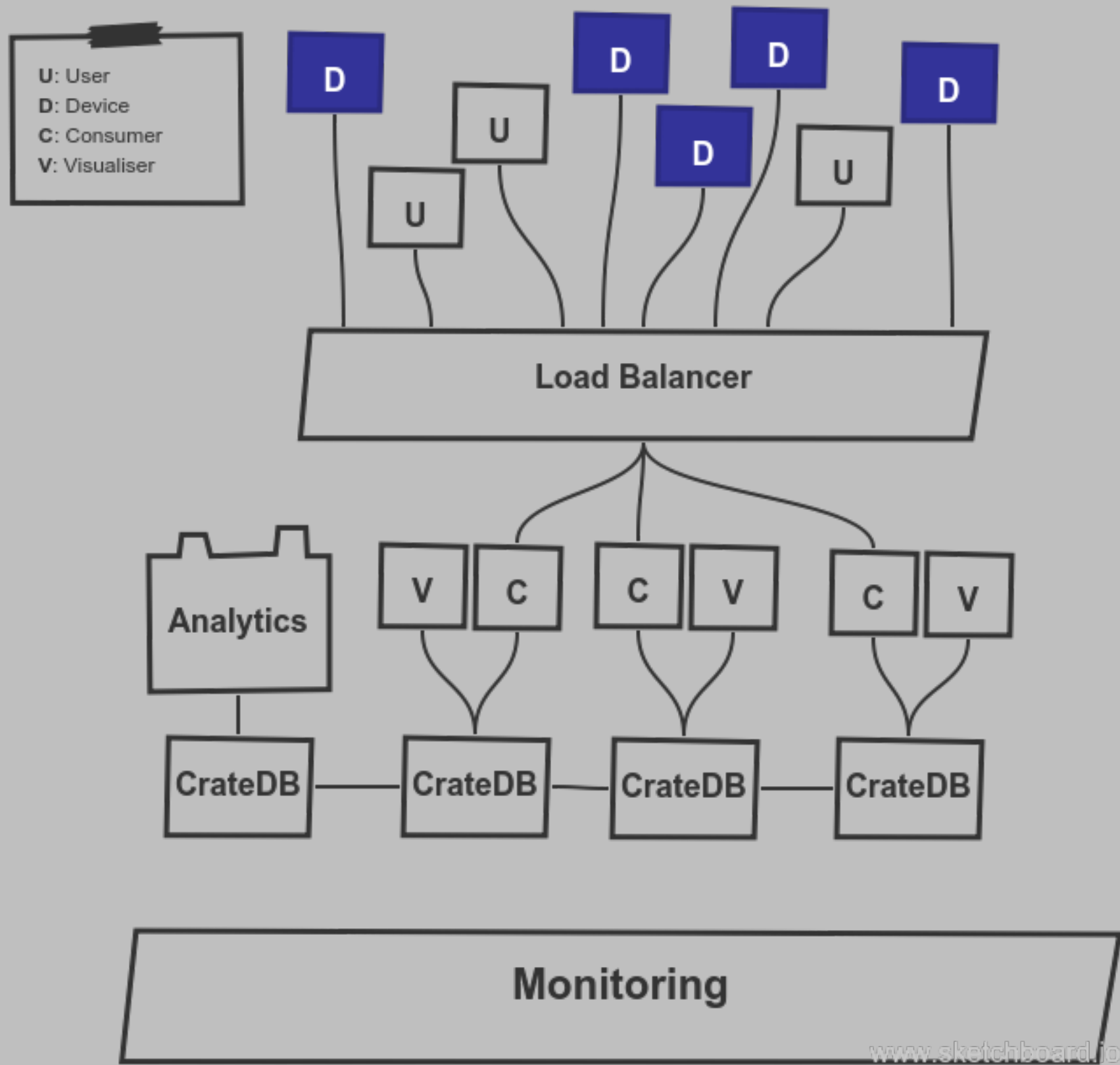
## INTELLIGENT SOLUTIONS FOR TOURIST ACCOMMODATIONS





# THE PROBLEMS

- Lot of network connections
- Relying on the message queue
- Duplicate data (queue + DBs)
- SQL DB does not scale horizontally as easy as the NoSql DB
- Access control
- Monitoring



# THE ADVANTAGES

- Simplified design (reduced technology stack)
- Consumers and Visualisers talk to the localhost
- Live reporting: Use ad-hoc queries on the production data
- No single point of failure
- No message queue
- DB isolation since it is accessible only from the localhost
- Scales horizontally almost linearly

LIVE DEMO  
and

<https://play.crate.io/#/>

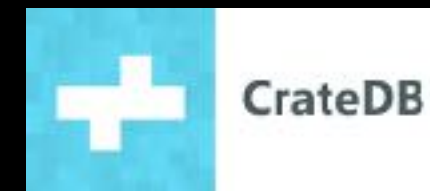
<https://gist.github.com/matriv/0ea8125c2a49a22b3533101f771c2063>







# Q & A



Presenter: Klearchos Klearchou  
GitHub: <https://github.com/klearchos/crate>  
LinkedIn: <https://www.linkedin.com/in/klearchosklearchou/>

