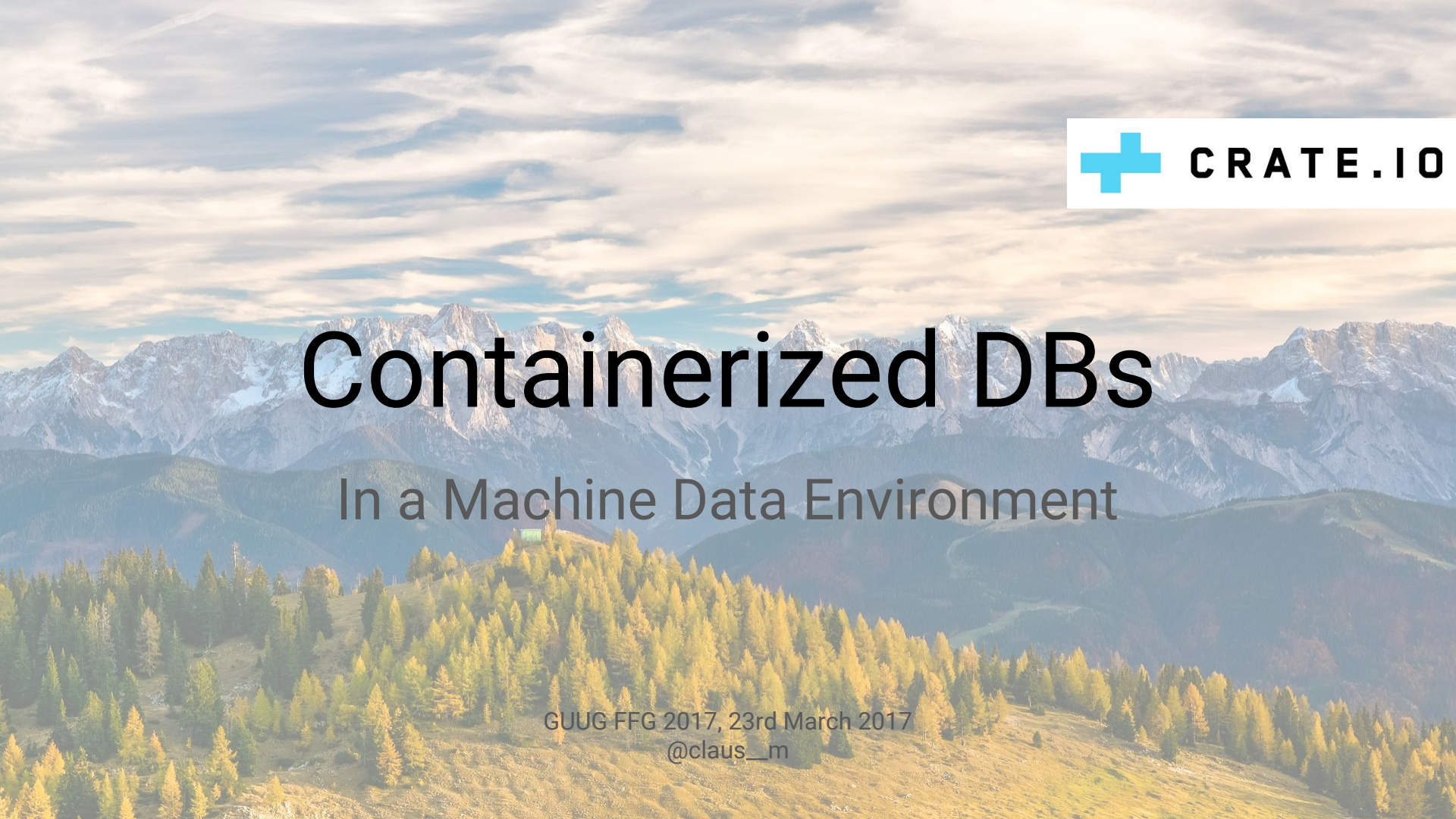


Containerized DBs

In a Machine Data Environment

GUUG FFG 2017, 23rd March 2017

@claus__m

About

~2yrs at Crate.io
DevRel/Field Engineering/Support/ Integrations/…

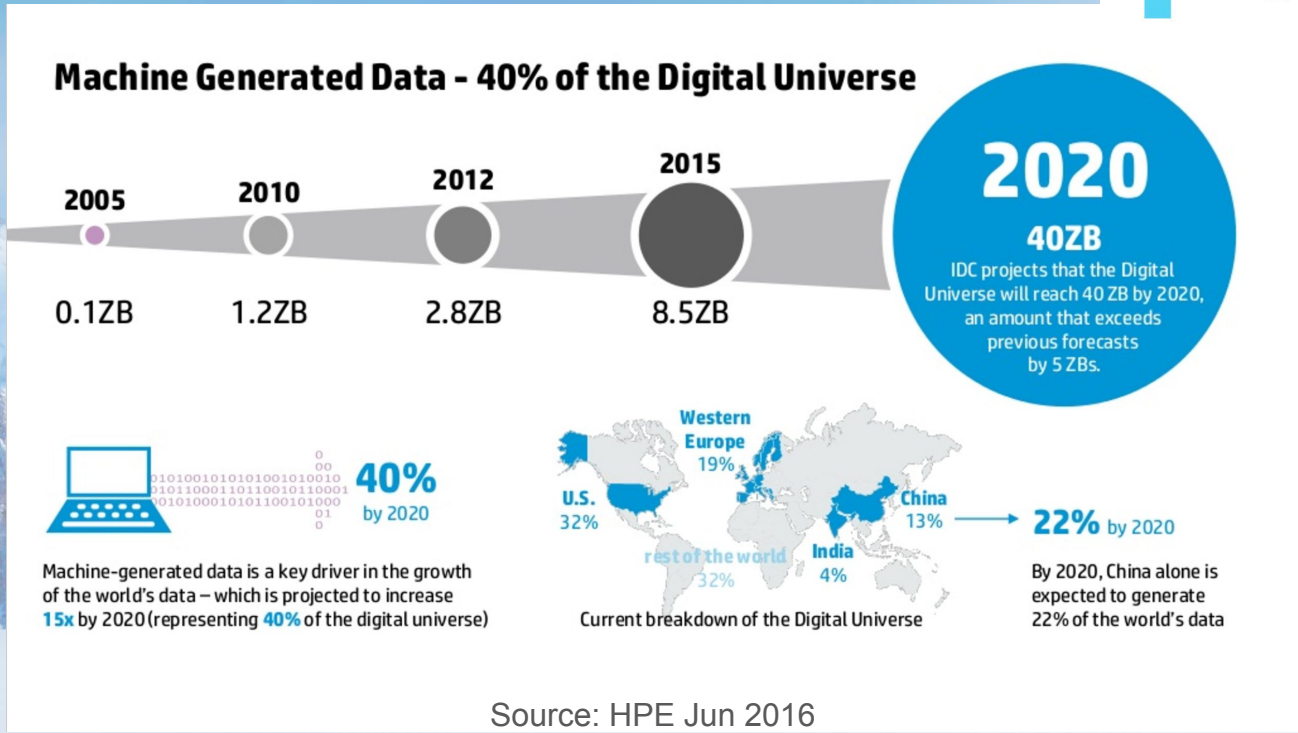
Crate.io
Founded in 2013, ~25 people and growing

Offices
San Francisco, Berlin, Dornbirn (AT)

Talk to me about
Rust, Raspberry Pis, Tech!


CRATE.IO
Machine Data
@claus__m

CRATE.IO



Source: HPE Jun 2016
http://www.slideshare.net/penumuru/harness-the-power-of-big-data-with-oracle-63438438/9

Machine Data Characteristics

Millions of data points/second
Streaming in from sensors, devices, logs, etc.

Data diversity
Structured & unstructured JSON, Blobs

Real-time query performance
Monitoring & alerting

Complex queries of big data volumes
With Terabytes of historic data

Growth
Adding sources often means exponential growth

CRATE.IO

Machine Data

Internet of Things
Sensors, cameras, ...

Wearables, Gadgets
Location data, interaction data, ...

Logs & Monitoring data
Component health monitoring, access logs, ...

Industry 4.0, Digitization
Production line insights, automation, ...

Vehicles
Location data, health data, ...

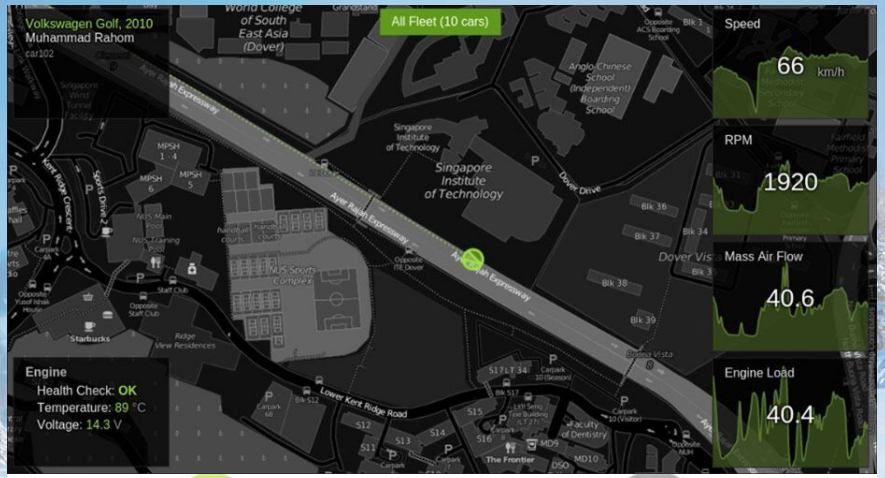
CRATE.IO

Clickdrive.io

Fleet management & vehicle tracking
Vehicle health and tracking data

High ingest rate
2,000 data points per car, per second

In-depth & real-time analysis
Predictive maintenance, accident reconstruction, route/driver efficiency



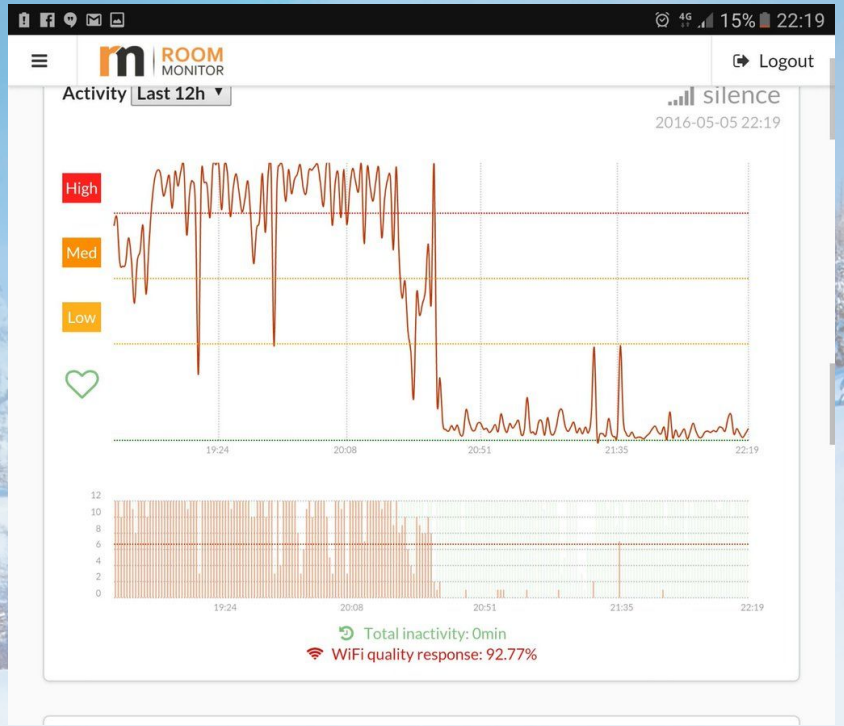
Roomonitor

Smart apartments

Monitoring & control climate, occupancy, noise, access

Better efficiency, safer environment

Alerts: AC/heating on with window open, noisy neighbors, ...

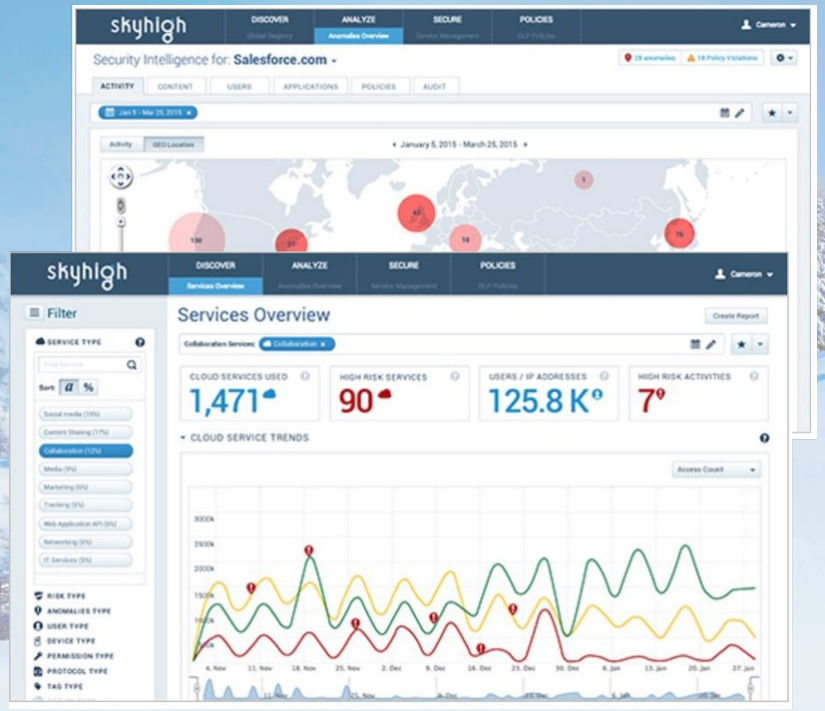


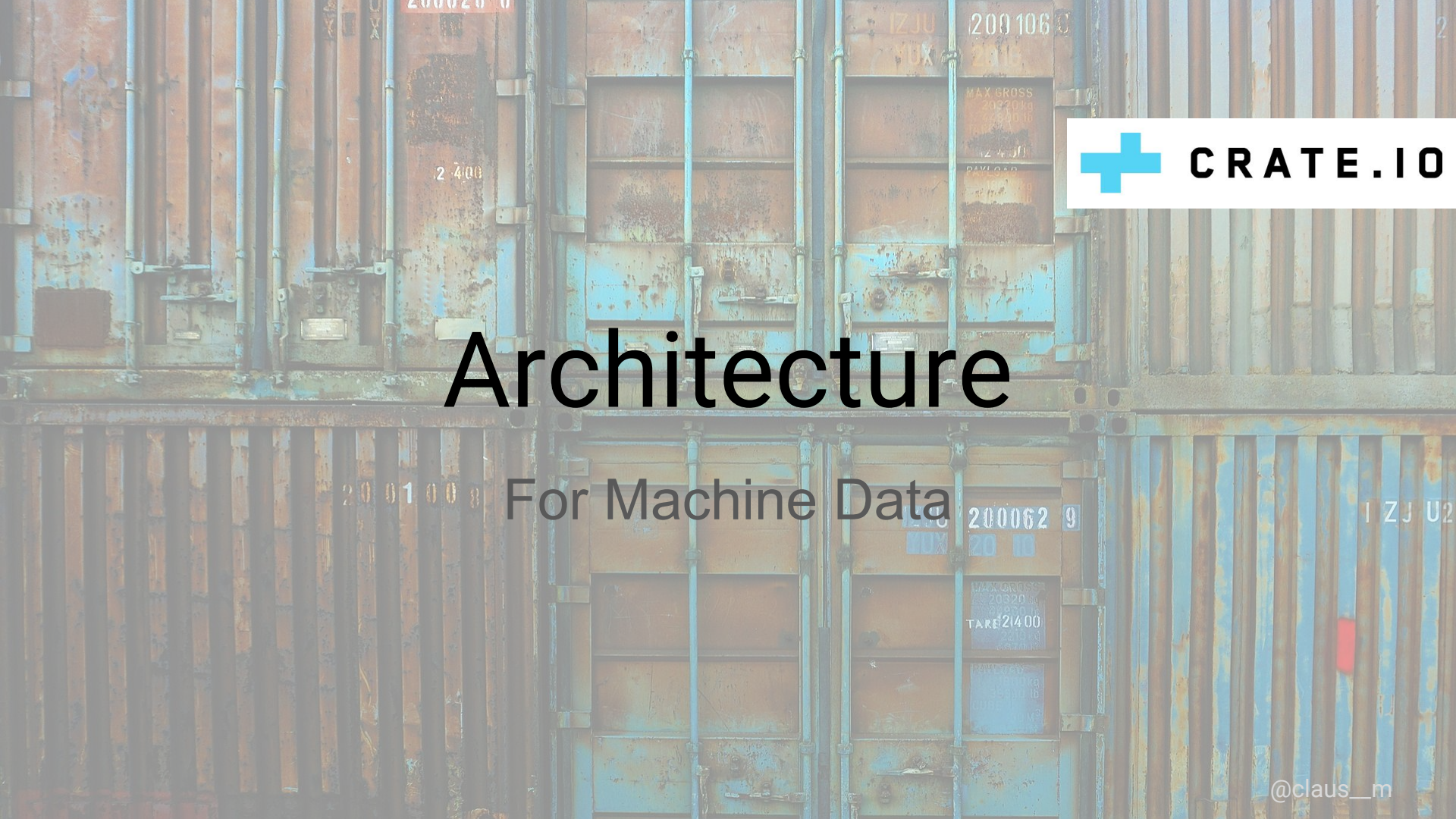
Skyhigh Networks

Cloud access security broker (CASB)
Access logging for cloud services

Large data volumes & ingest
Billions of events per day from 600+ customers, 10s of thousands of concurrent TCP connections

Machine data is the fingerprint of fraud
Unsupervised learning to find anomalies





Architecture

For Machine Data

@claus__m

Microservices

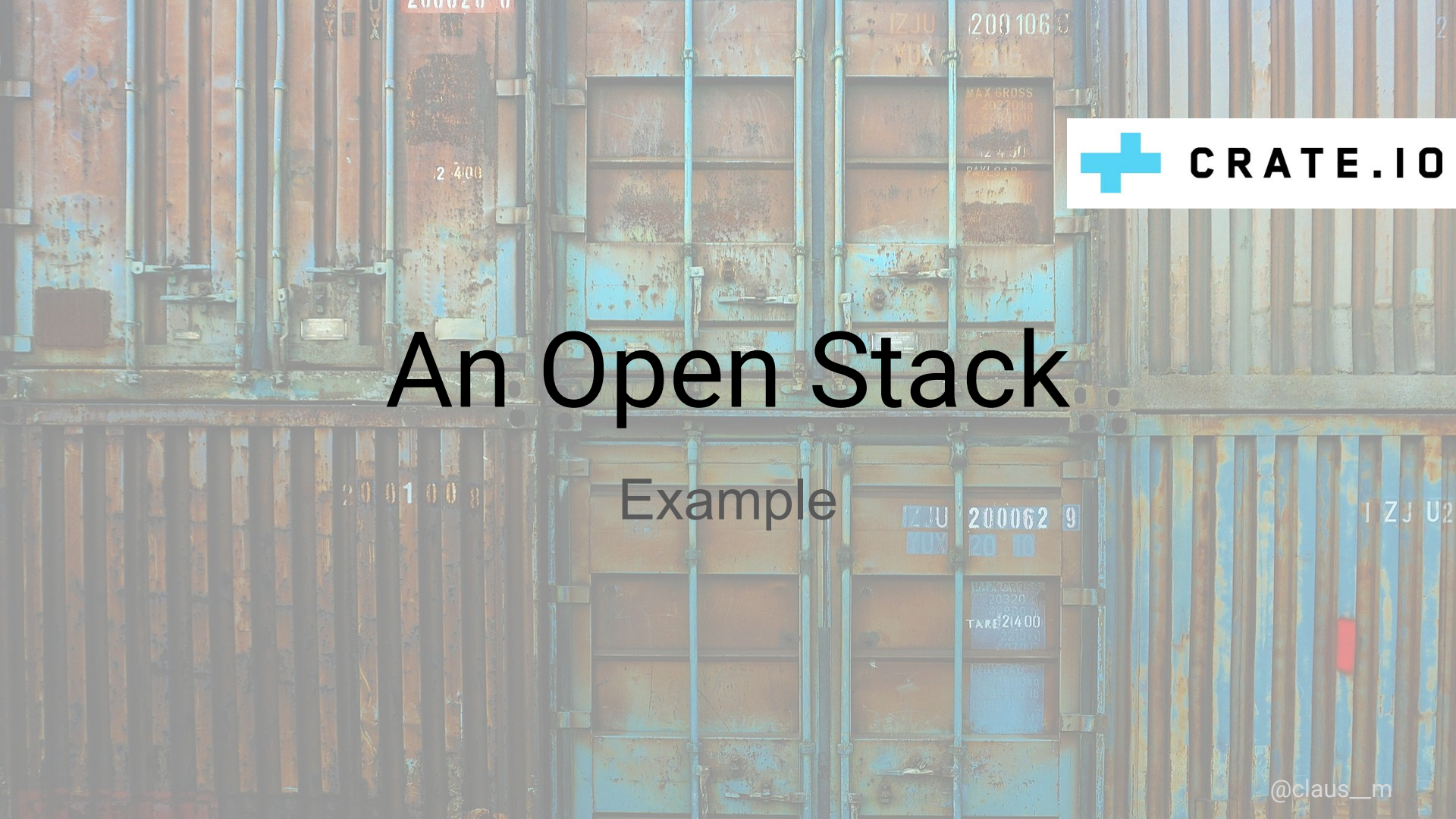
Containers
Isolation by default

Flexibility
Building blocks

Horizontally scalable
Mostly

Stateful containers
Databases?

CRATE.IO



An Open Stack

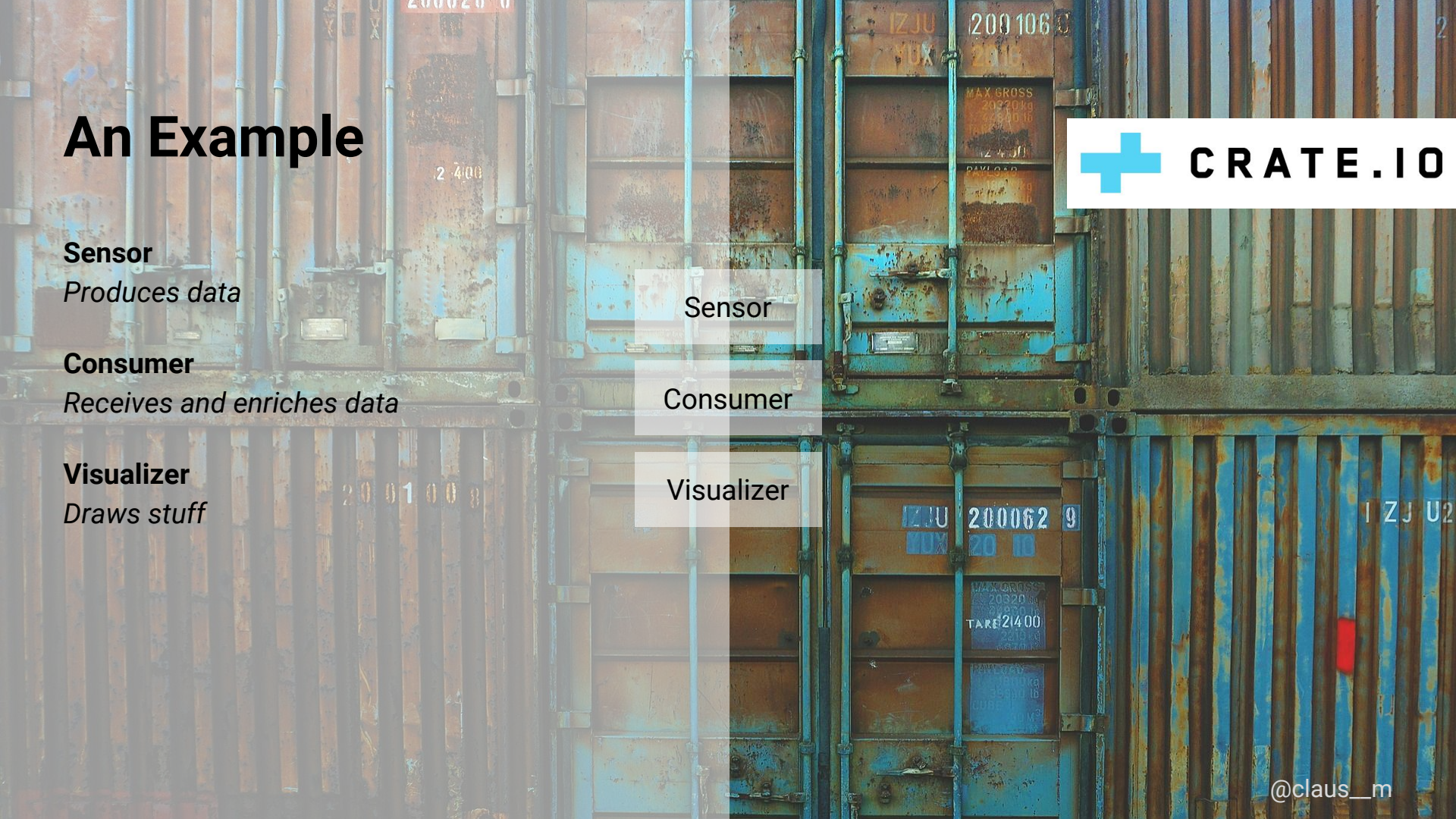
Example

An Example

Sensor
Produces data

Consumer
Receives and enriches data

Visualizer
Draws stuff

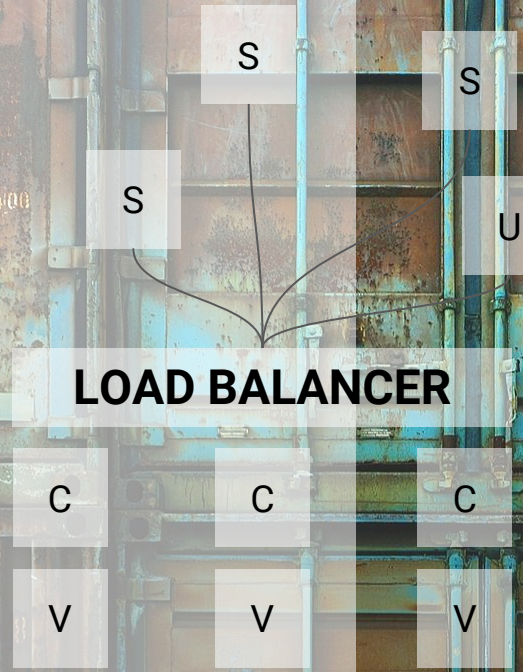


Deploy!

Load balancer
For TLS, reverse proxying, load balancing

High availability
3 instances

A few sensors
One user to actually use it



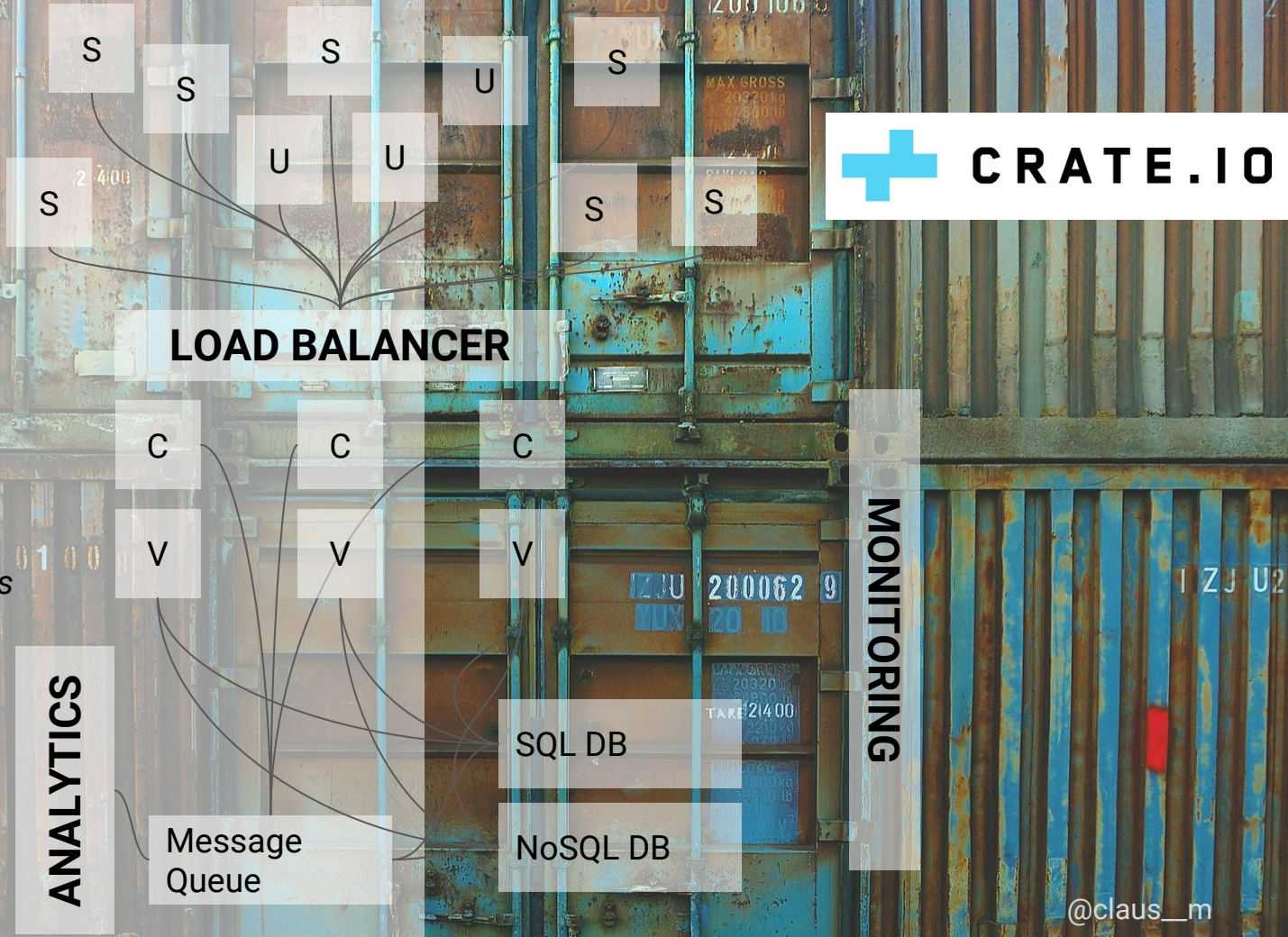
CRATE.IO

Go Live

More users!
More sensors and users

Data storage
Slow and fast

Monitoring & Analytics
Two different subsystems



But ...

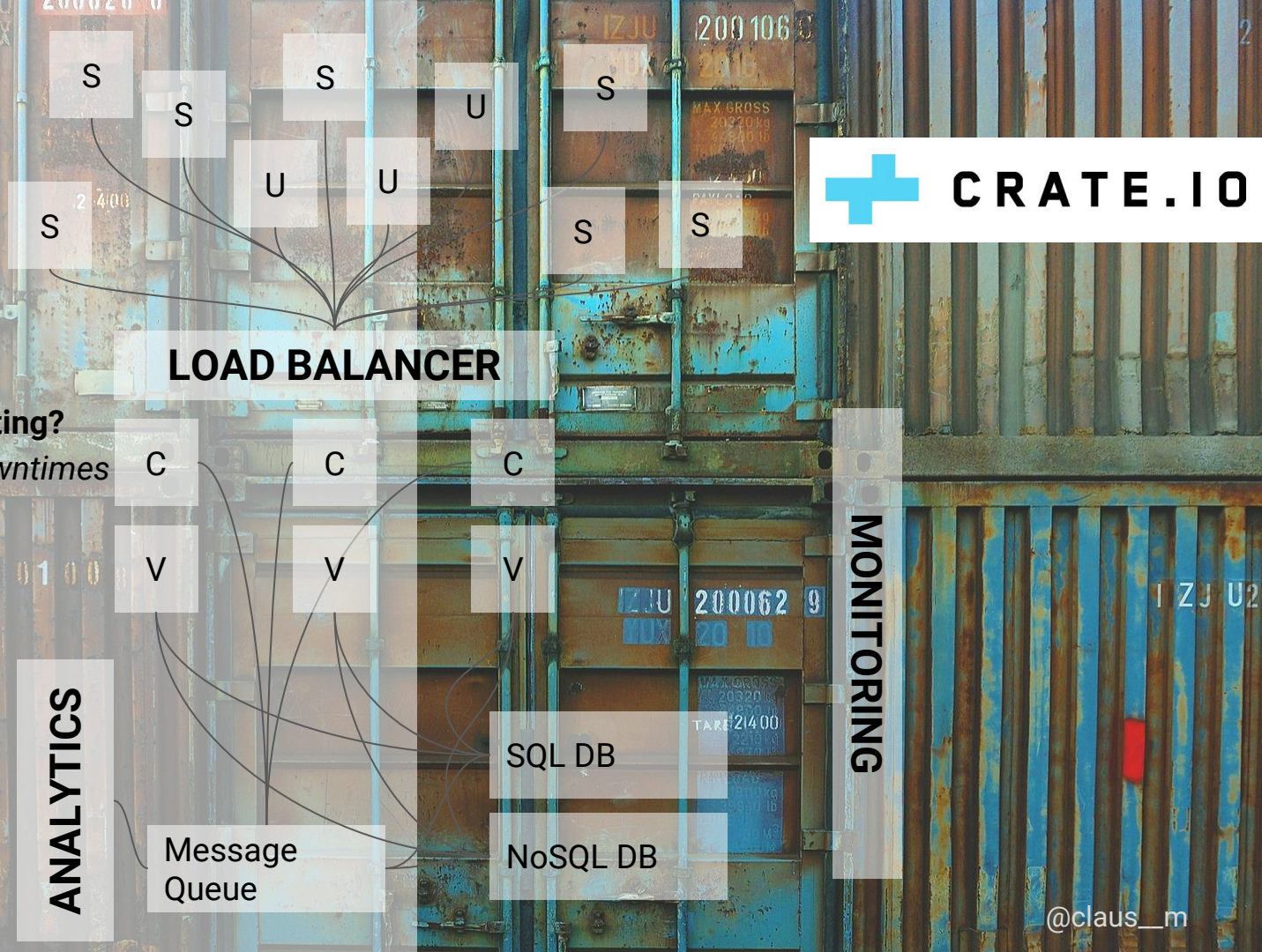
Even more users?
Horizontal scaling?

Maintenance & bug hunting?
Mostly via scheduled downtimes

Reporting?
Kafka? Elasticsearch?

Security?
Access control?

Expertise?
Knowledge transfer?



CRATE.IO

A Database for Microservices

Shared nothing

Replication

Sane defaults

Resilient

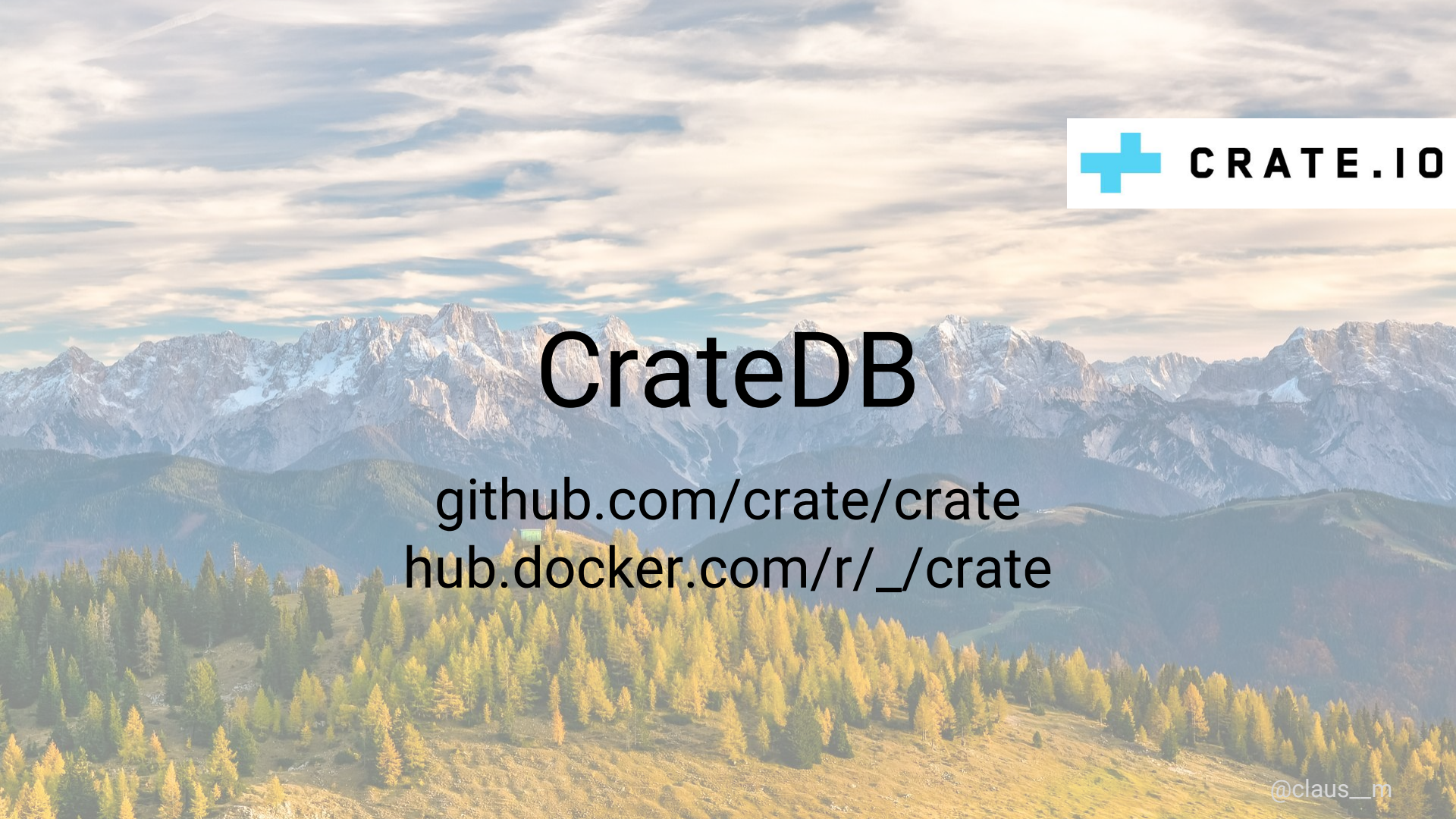
Cross-functional

CRATE.IO

@claus__m


CRATE.IO
Another DB?
Yay!
@claus__m

CRATE.IO


CRATE.IO
CrateDB
github.com/crate/crate
hub.docker.com/r/_/crate
@claus__m


CRATE.IO
Solomon Hykes
@solomonstre
Following
CrateDB 1.0 today. Very cool project built by amazing human beings
@claus__m

CrateDB

Shared nothing

Partitioning & auto-sharding

Replication

(Almost) Zero config

Multi model: Structured & unstructured

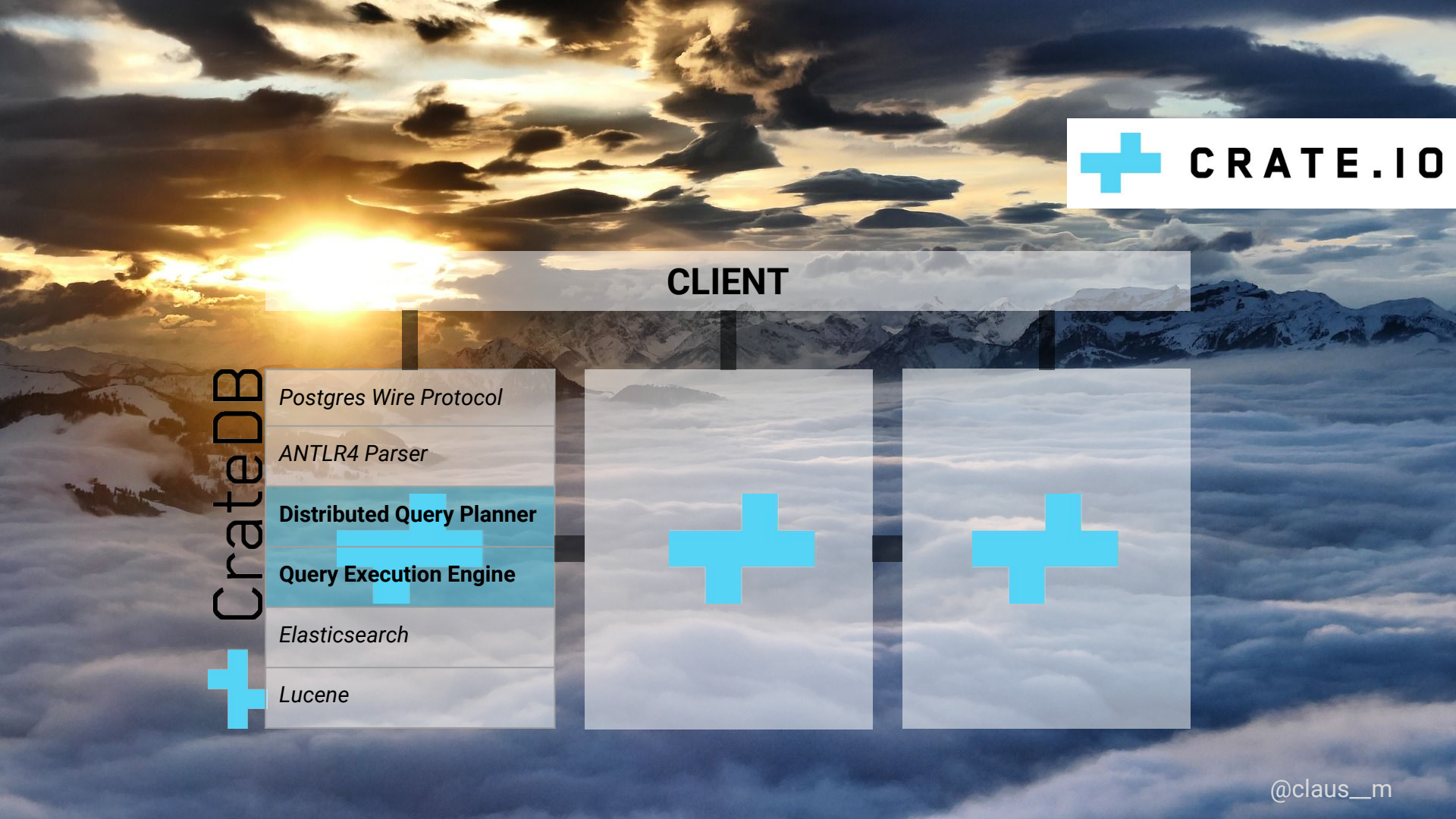
SQL

CrateDB Fundamentals

Disk-based index with in-memory caching
Fast and efficient OS caching

Shards: Units of data
Concurrency by distributing shards

Distributed query execution engine
“Push down” queries


CRATE.IO
CLIENT
CrateDB
Postgres Wire Protocol
ANTLR4 Parser
Distributed Query Planner
Query Execution Engine
Elasticsearch
Lucene
@claus__m

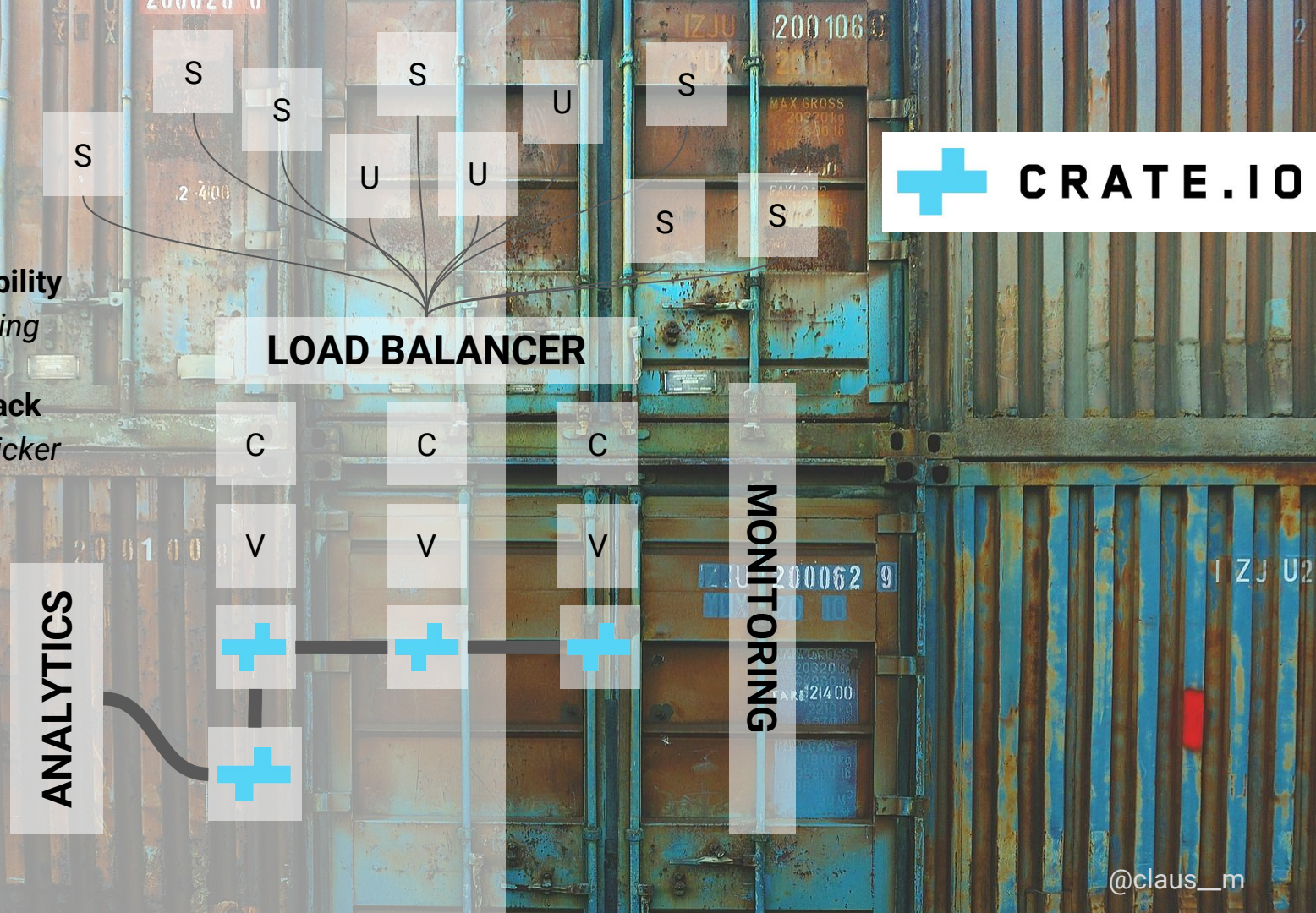
A better setup!

Horizontal scalability
Scale out everything

Reduced tech stack
Get to know it quicker

Live reporting
Use ad-hoc queries on production data

Flexibility
Schema Evolution not required



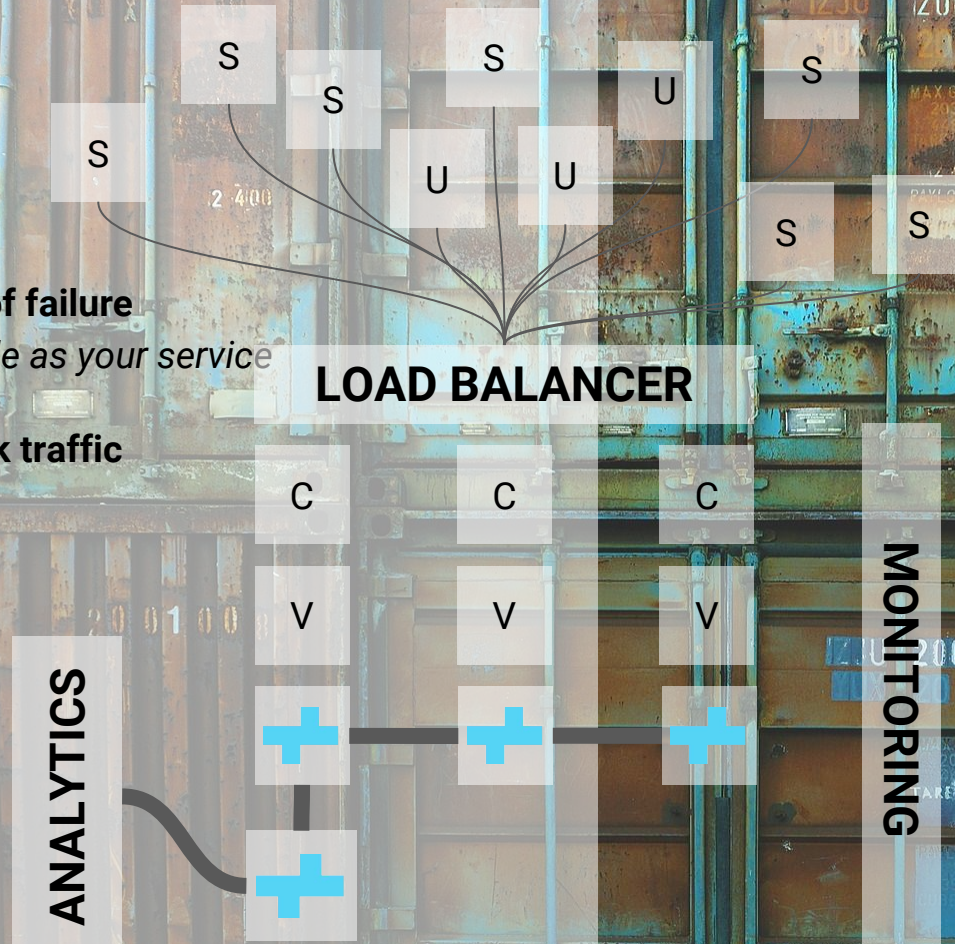
A better setup!

No single point of failure
As highly available as your service

Reduced network traffic
Better reliability

No queue
Work with real data

DB isolation
Accessible only from the host



CRATE.IO

Live Demo

Docker Swarm
Orchestration across platforms

Eden Server (Rust!)
RESTful web service

Eden Client (Rust!)
ARM application for reading temperature data from BMP180

Grafana
To draw up a nice dashboard

CRATE.IO


CRATE.IO
Demo Time!
@claus__m

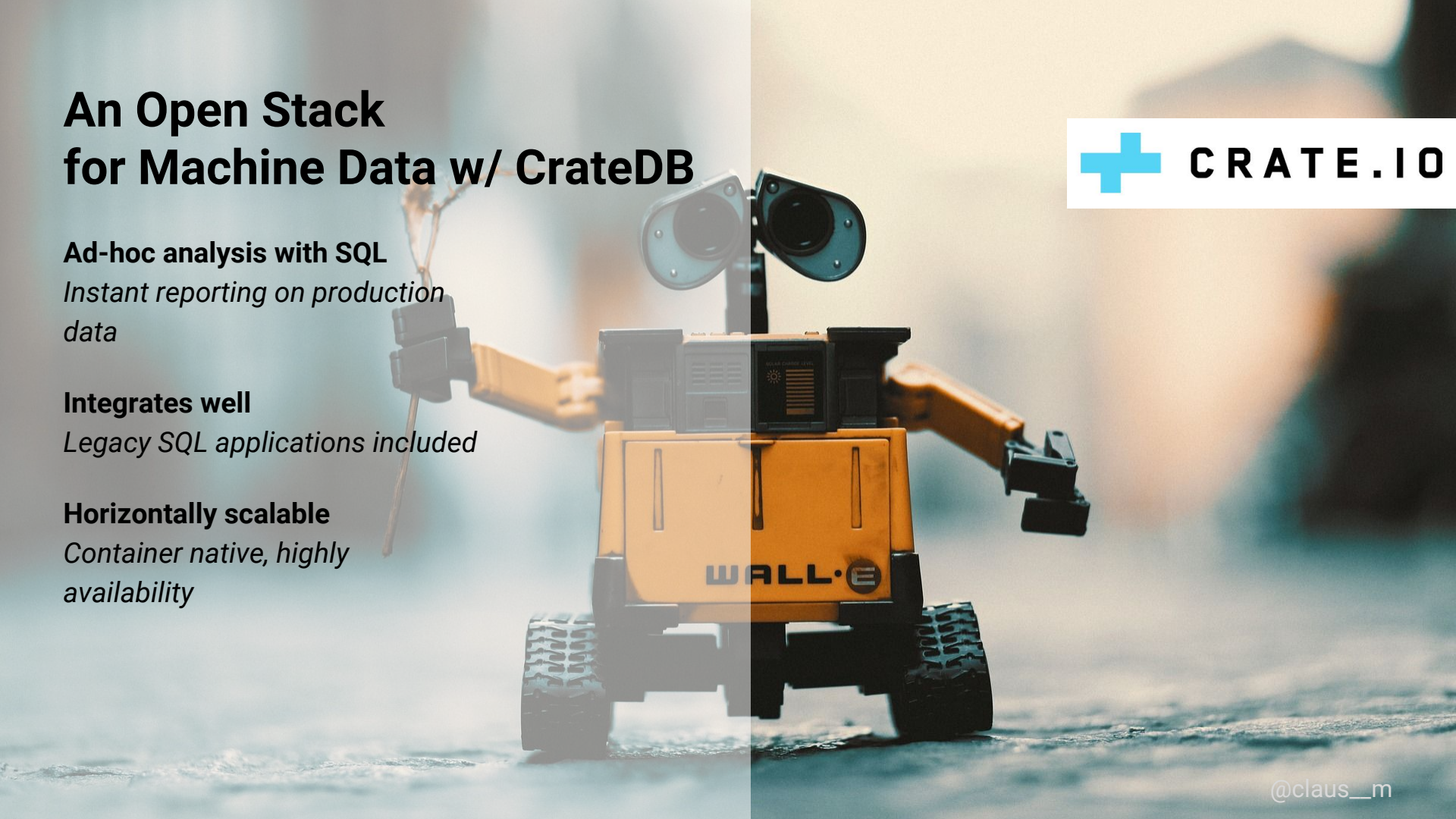
An Open Stack for Machine Data w/ CrateDB

CRATE.IO

Ad-hoc analysis with SQL
Instant reporting on production data

Integrates well
Legacy SQL applications included

Horizontally scalable
Container native, highly availability



@claus__m

Links

https://github.com/celaus

https://github.com/crate

https://hub.docker.com/r/_/crate

https://crate.io

Follow us on twitter
@crateio @*claus__m*

Thanks!