

CRATE.IO

USING GRAFANA AND CRATEDB

Processing Your Machine Data

Paul Adams | Engineering Manager



WHAT IS CRATEDB?

CrateDB is an open source, distributed SQL database with integrated search that makes it simple to store and analyze massive amounts of structured and unstructured data in real-time.

We love data:

- Time Series
- Geospatial
- BLOBS
- (Un)structured

A wide-angle photograph of a mountain range. In the foreground, there are dark, rocky mountain slopes. Above them, several peaks are covered in white snow. The sky is filled with soft, grey clouds, creating a serene atmosphere.

HOW DID WE GET HERE?

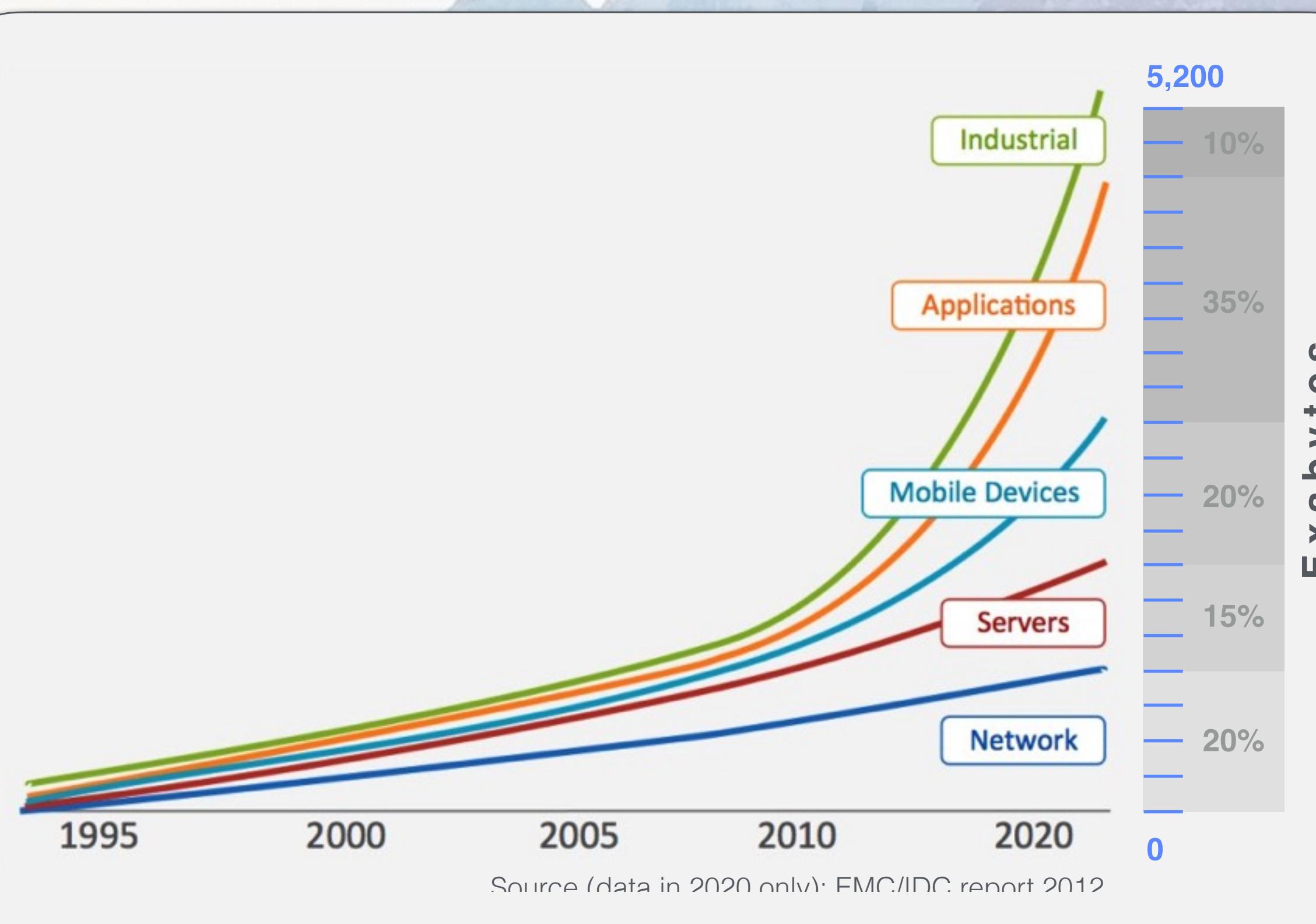
2014 FOUNDING MISSION: A SQL DB FOR THE MACHINE DATA ERA

Market need:

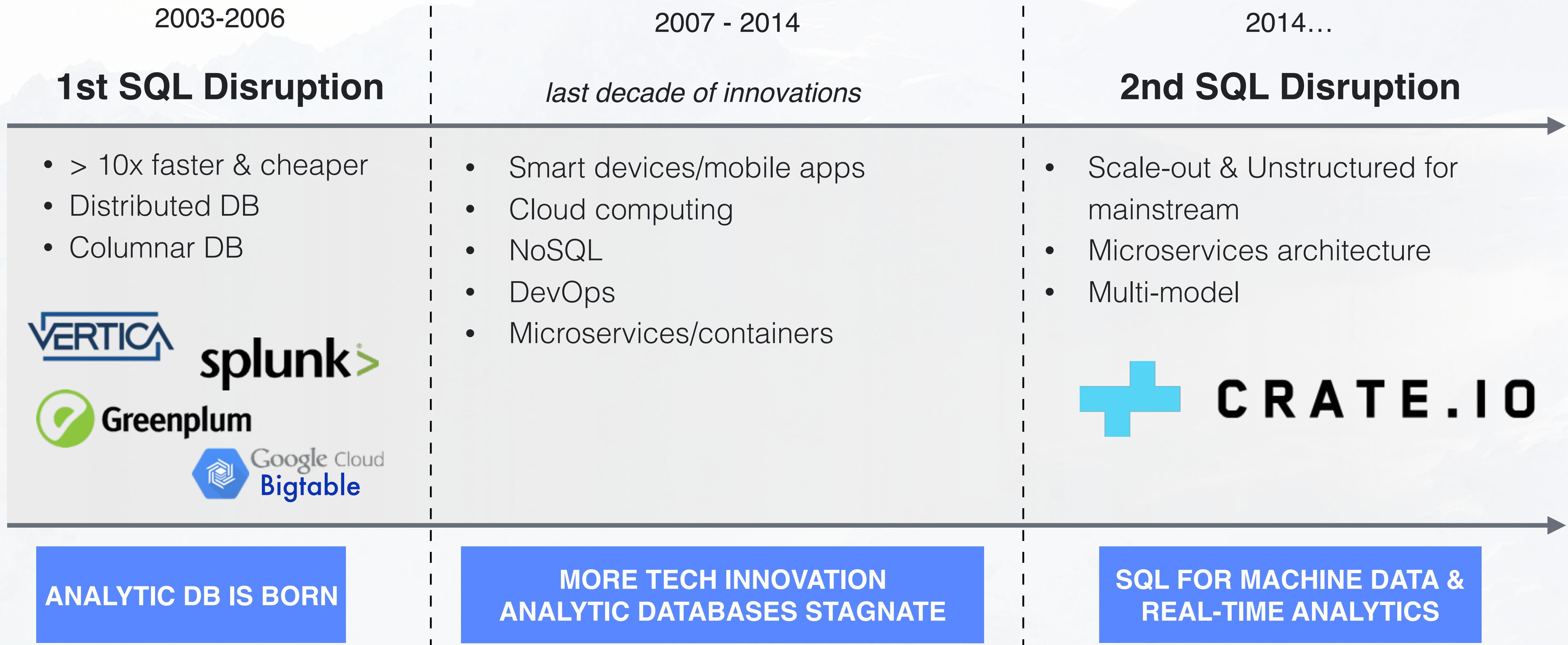
- A machine data analytics database for the SQL mainstream

Key design elements:

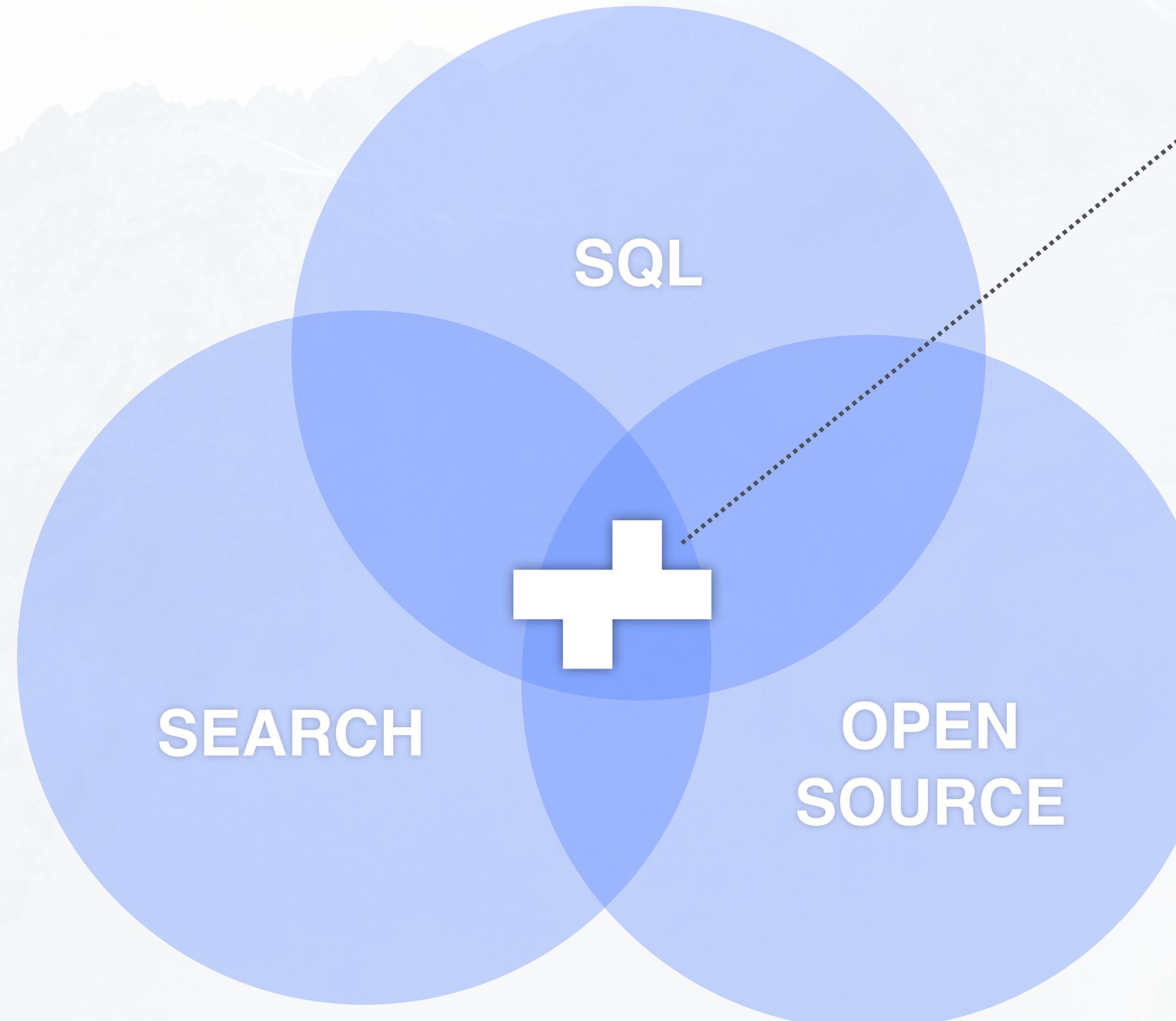
- Millions of data points/second
- Real time SQL query performance
- Time series, geo, predictive, search...
- Ad-hoc & aggregations!
- Any structured or unstructured data
- Simple scaling
- SQL access



BUILT ON A DECADE OF RECENT INNOVATIONS...



CRATEDB COMBINES BEST OF ALL WORLDS



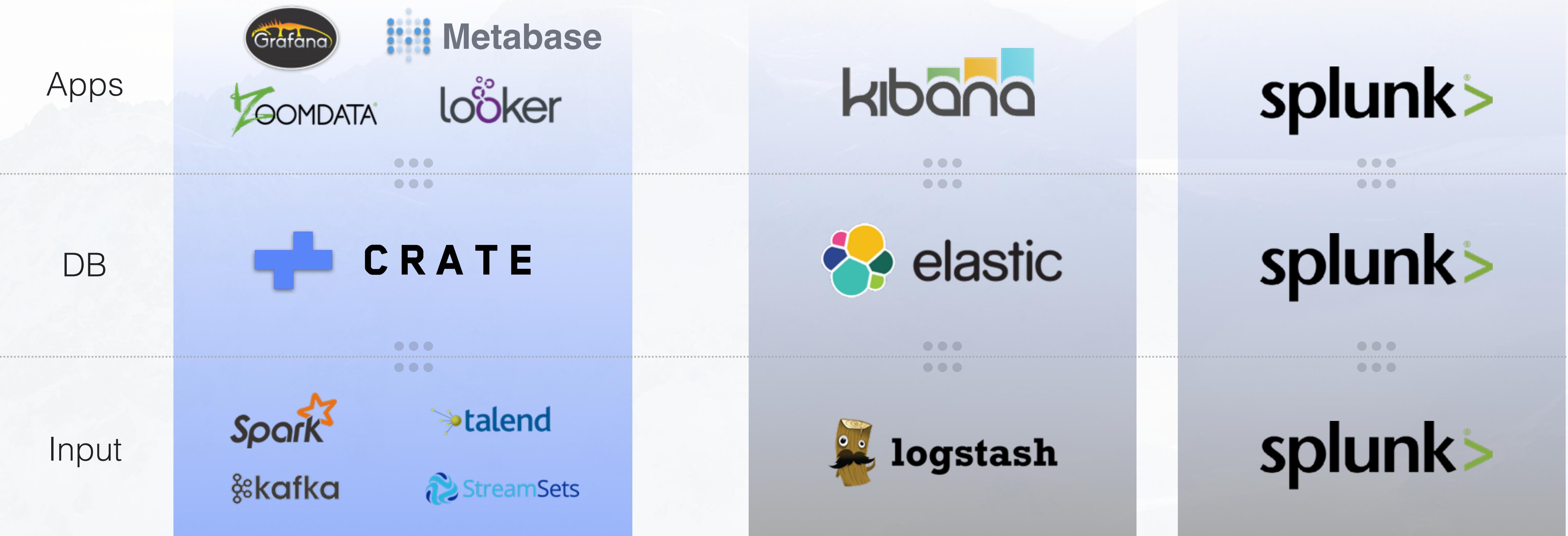
Perfect for workloads with:

- Multi-structured data
- High velocity INSERT
- Fast queries
 - Ad hoc
 - Time series
 - Geospatial
 - Aggregates
- Scale-out architecture
- Fault-tolerant
- BI integration

All easy enough for SQL mainstream to operate

Enabled by distributed query and micro services architecture innovations

OPEN MACHINE DATA STACK



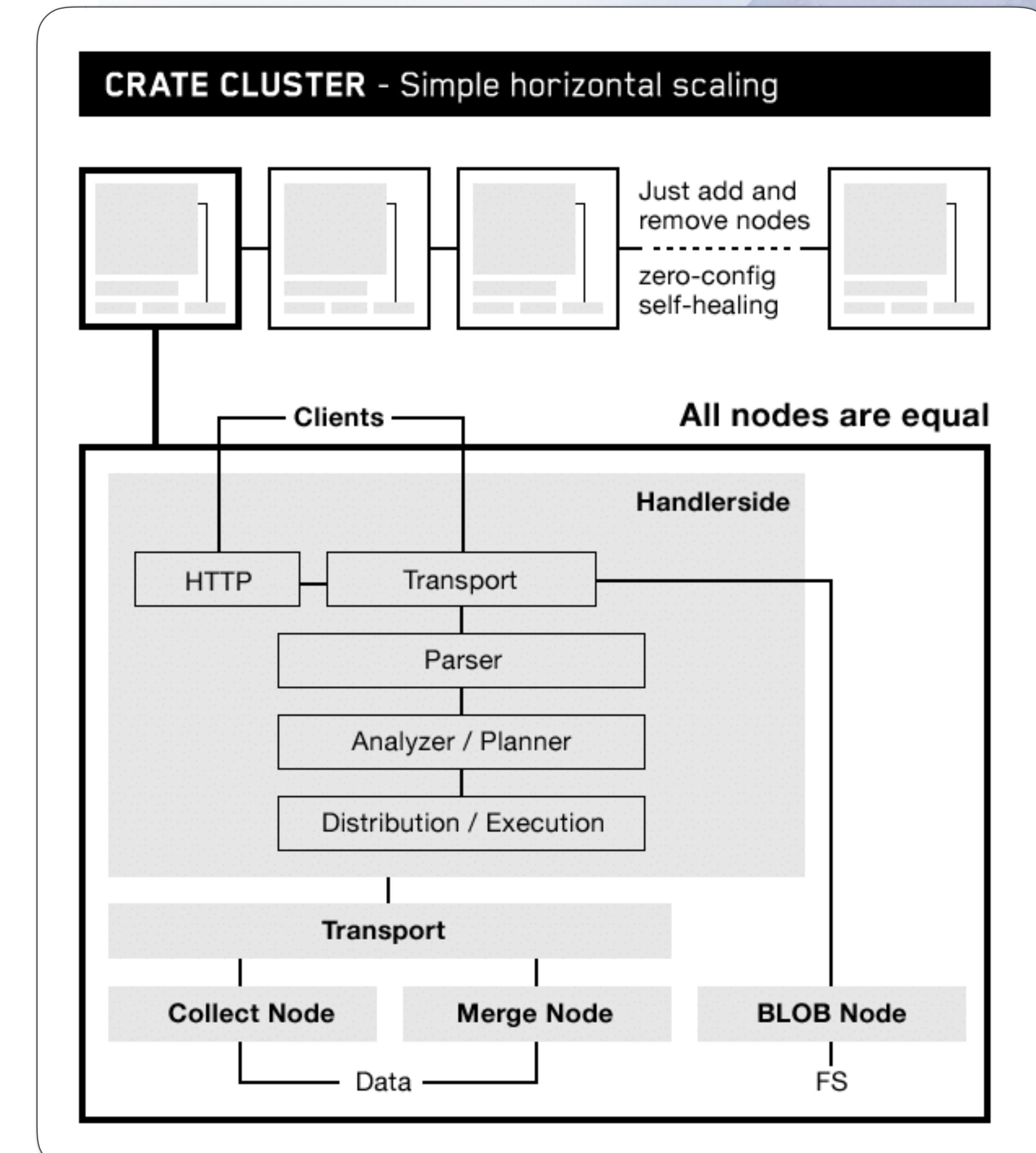
- *Integrates easily*
- *Low learning curve*
- *Greatest flexibility*
- *No lock in*



CRATE DB FEATURES & ARCHITECTURE

CRATEDB KEY IDEAS

- Distributed SQL for scale out
 - Simple scalability
 - Masterless, shared-nothing microservices architecture
 - Auto-sharding & partitioning
- Realtime search & aggregations
- Columnar, multi model
- Dynamic schema
- Timeseries, Geospatial support
- In-memory speed



SQL FEATURES:

- ANSI SQL (subset with extensions)
- JOINS (fully distributed)
- Arrays and nested objects
- Different types
- Information Schema
- Cluster/Node state exposed via tables
- Partitioned tables
- Geospatial support
- Full text search
- Powerful text processing
- Subselects (en route)
- Streaming support
- Common relational Operators:
 - Projection
 - Grouping (incl. HAVING)
 - Aggregations
 - Sorting
 - Limit/Offset
 - WHERE-clause
- Import/Export

GRAPHICAL ADMIN UI + SHELL

CLUSTER: CRATE VERSION: 0.54.7 NODES: 8 HEALTH: DATA CHECKS LOAD: 2.51 / 2.26 / 1.41

?

!

Cluster: crate

good 100.0% 100.0% 3.3 Billion 0 0

Health Replicated Data Available Data Total Records Underrepl. Records Unavail. Records

Cluster Load

cr> select landmark, count(*) from "nyc-2 desc limit 10;

Landmark	count(*)
J F K AIRPORT	5194
LA GUARDIA AIRPORT	4829
CENTRAL PARK	1537
PENNSYLVANIA STATION	1064
PORT AUTH 42 STREET	703
GRAND CENTRAL TERM	662
PROSPECT PARK	509
YANKEE STADIUM	273
CITY HALL	233
SI FERRY TERMINAL	226

SELECT 10 rows in set (0.060 sec)

Load 1 Load 5 Load 15

?

A wide-angle photograph of a mountain range. In the foreground, there are dark, rocky mountain slopes with patches of snow. In the background, several peaks rise against a sky filled with soft, white clouds. A winding road or path is visible on one of the upper ridges.

NYC 311 + CABS

WHAT IS 311?

- Source of government information and non-emergency services:
 - Heating
 - Sanitation
 - Potholes
- ~24million incidents reported since 2009
 - 100+million calls
 - Data published monthly!



```
cr> select incident_address, count(*) from "nyc-311" where complaint_type  
incident_address is not null group by 1 order by 2 desc limit 10;  
+-----+-----+  
| incident_address | count(*) |  
+-----+-----+  
| 2172 2ND AVENUE | 305 |  
| 259A KINGSTON AVENUE | 179 |  
| 37-14 73 STREET | 99 |  
| 96 8 AVENUE | 91 |  
| 2655 RICHMOND AVENUE | 81 |  
| 518 WEST 27TH STREET | 79 |  
| 2161A WHITE PLAINS ROAD | 78 |  
| 3340 BROADWAY | 76 |  
| 90-15 0 |  
| 37-11 7 |  
+-----+  
SELECT 10
```

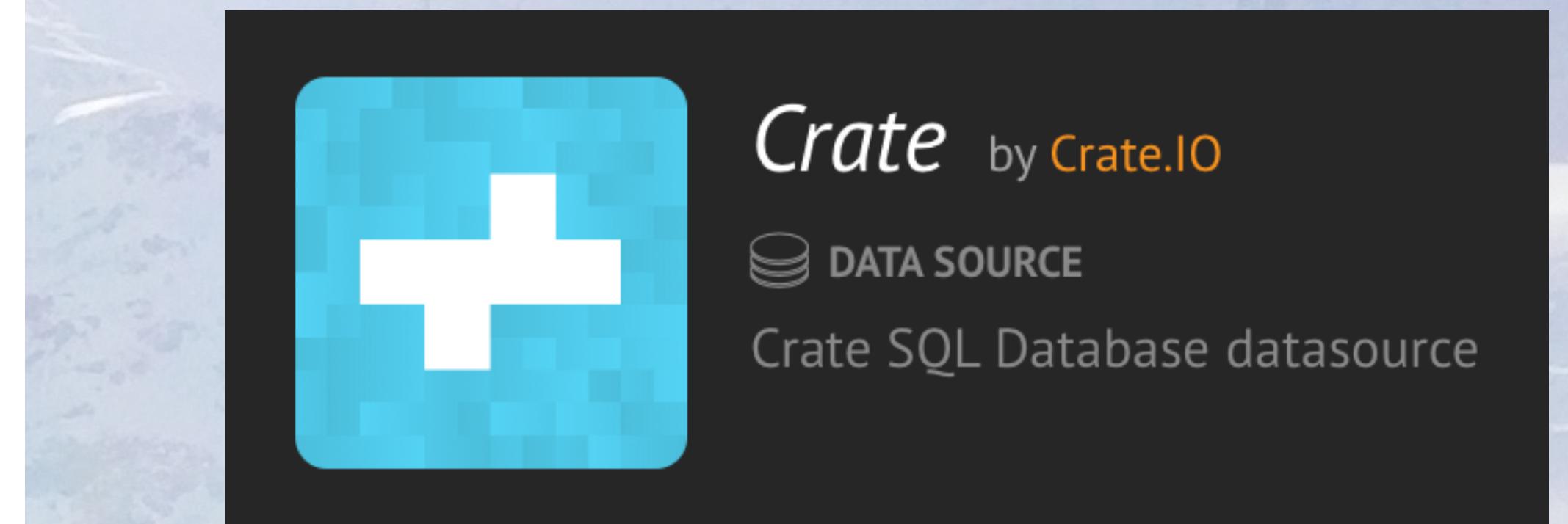


The background of the image is a wide-angle photograph of a mountain range. In the foreground, there are dark, rocky mountain peaks. Above them, the middle ground shows more rugged terrain with patches of snow and ice. The background features a range of mountains with prominent peaks, some of which are capped with snow. The sky is filled with soft, white clouds, suggesting either dawn or dusk. A large, solid blue rectangular box is positioned in the center-left area of the image, containing the text.

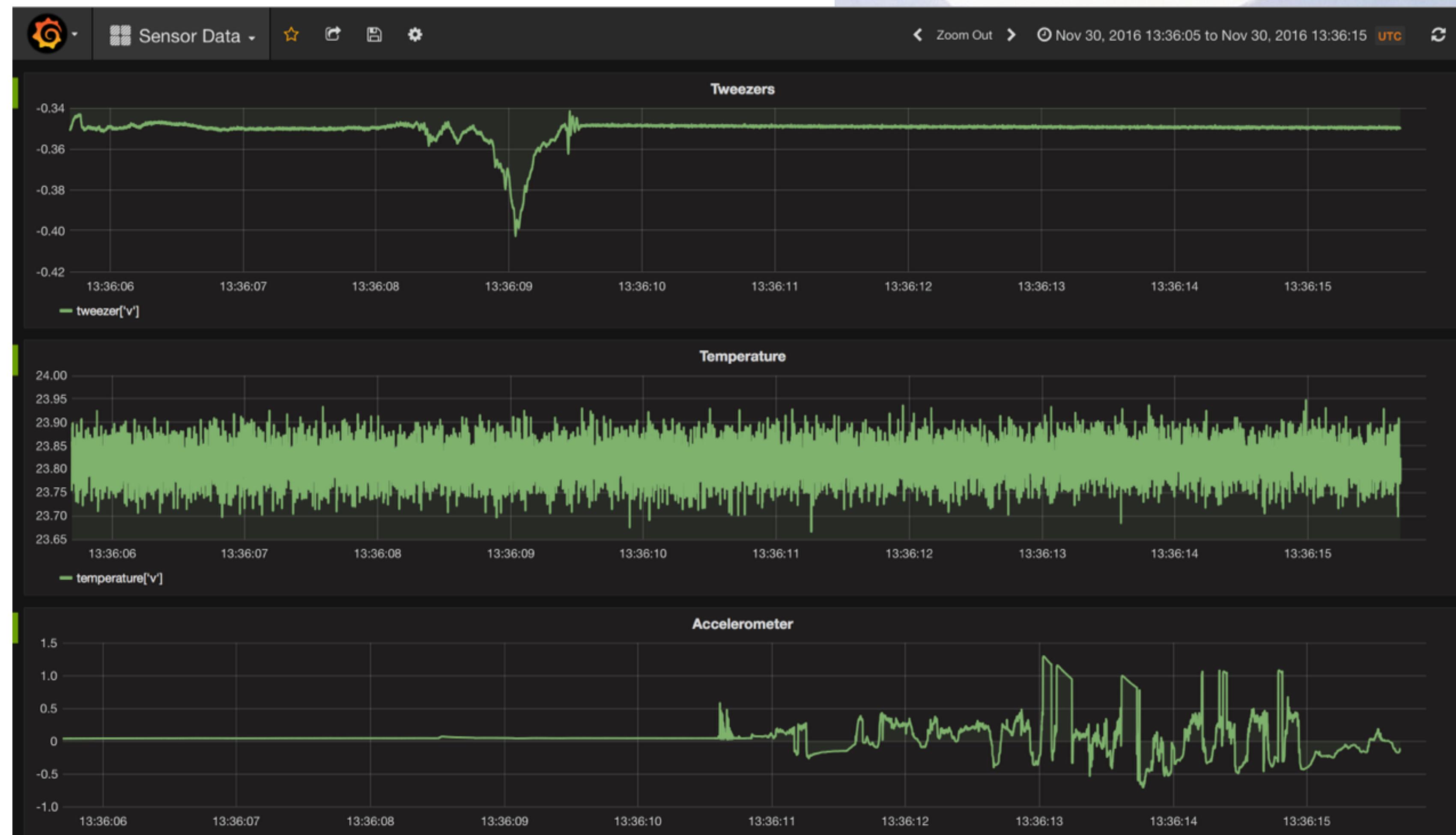
CRATEDB + GRAFANA

CREATEDB DATA SOURCE

- Version 0.2.0 now available!
 - Implementation by raintank GrafanaLabs (thanks!)
 - Feedback appreciated
- Support for all of CrateDB
 - Data types
 - Scalar functions: avg, min, max, abs etc



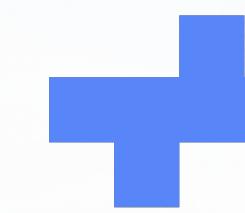
TIME SERIES MACHINE DATA



GANTNER Q.BRIXX + STATION

- Data Acquisition Device
 - Readings to 100khz
 - Output direct to CrateDB
 - Designed for industrial and test engineering environments
 - Voltage, Current, Temperature, Acceleration...
- The kind of thing the TSA *loves* to find in your hand luggage





CRATE.IO

THANK YOU

paul@crate.io

@therealpadams