Streaming in a World of Legacy Applications

David Brimley <David@hazelcast.com>
Vladimir Schreiner <Vladimir@hazelcast.com>

What is a Legacy System?

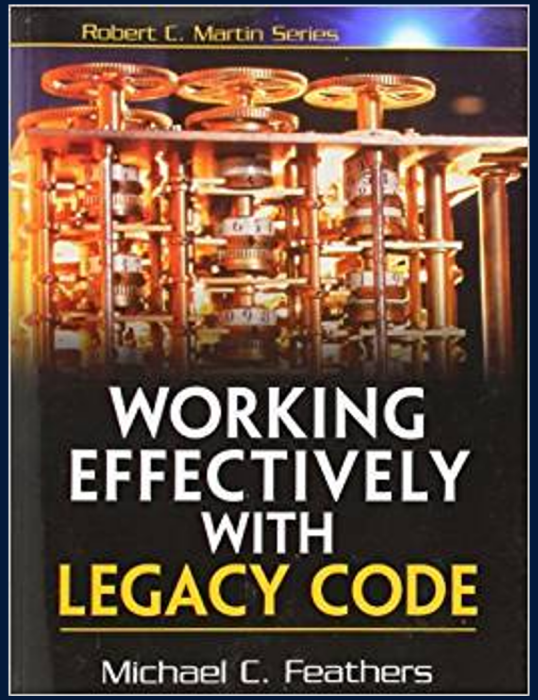
No Documentation.

The original devs are gone.

hazelcast

Zero Tests

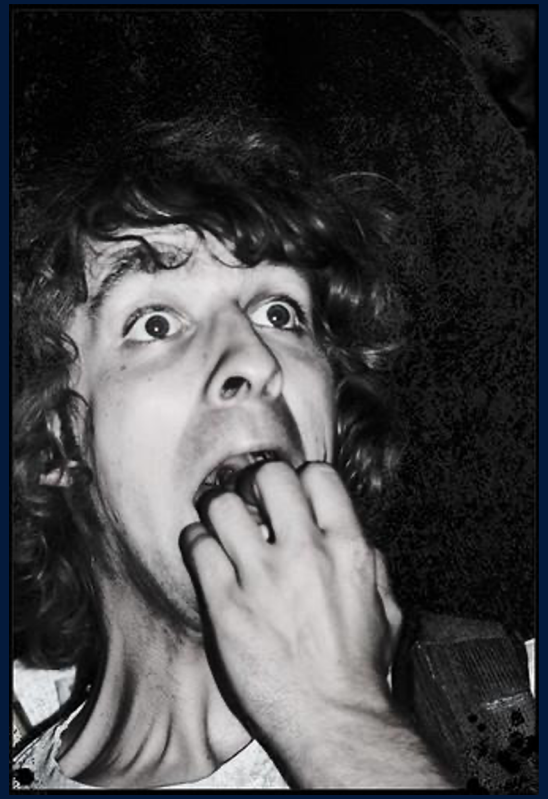
Time


Robert C. Martin Series
WORKING EFFECTIVELY WITH LEGACY CODE
Michael C. Feathers

THE MONEY MAKER!

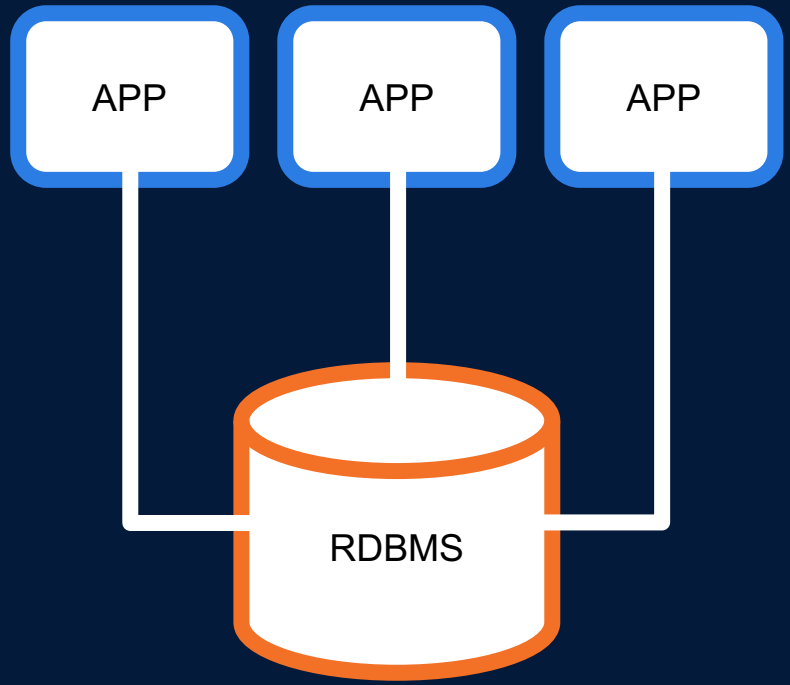
$ £ € ¥ ₽ Rs

Wayyyyyyyyyyyyyy to scary to risk your career changing it.

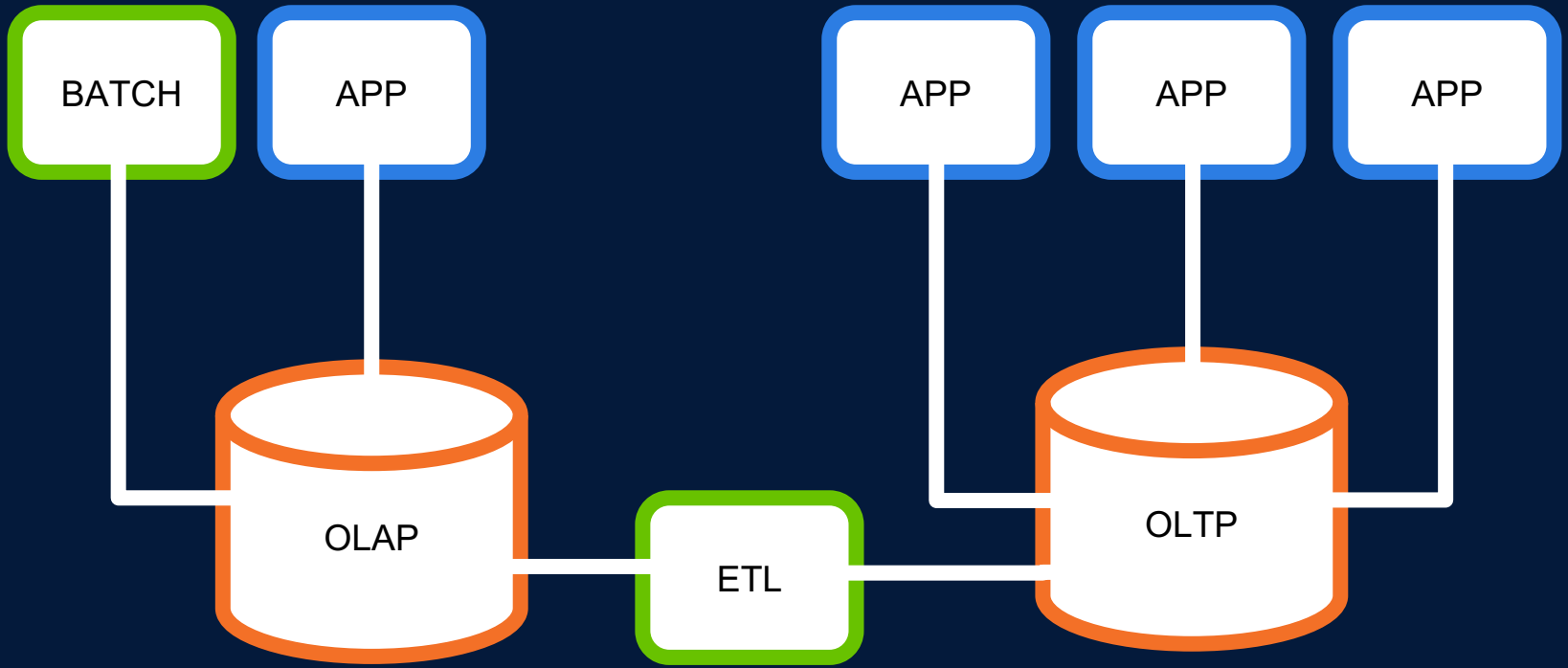


Usually built around a core Relational Database.

Typical Legacy Architecture



hazelcast


BATCH
APP
APP
APP
APP
OLAP
ETL
OLTP

Reasons for Change.

1.
Rooted in the present.

Current Database.

A **current database** is a conventional database that stores data that is valid *now*. For example, if a user inserts "John Smith" into the Staff table of a current database, this asserts that the fact is valid *now* and until it is subsequently deleted. By contrast, a temporal database qualifies each row with a valid time *stamp*, valid time *period* or valid time *interval*. For example, we can assert the fact that "John Smith" was a member of staff during the period 1 June 2001 and *now*.

https://en.wikipedia.org/wiki/Current_database

Ignore Temporal Value of Data

!Predictive Analysis

What if...

hazelcast

- Time Period Definitions
  - Start & End Times
- Temporal Primary Keys
- Temporal Referential Integrity
- Temporal Predicates
  - Contains
  - Overlaps
  - Equals
  - Precedes
  - Succeeds
  - Immediately Precedes
  - Immediately Succeeds
- Time Slice Queries

SQL:2011

2.
Answer Questions Now.

hazelcast

Store then Process

hazelcast

Process then Store

hazelcast

3.
Distributing Events

hazelcast

Most RDBMS are passive stores

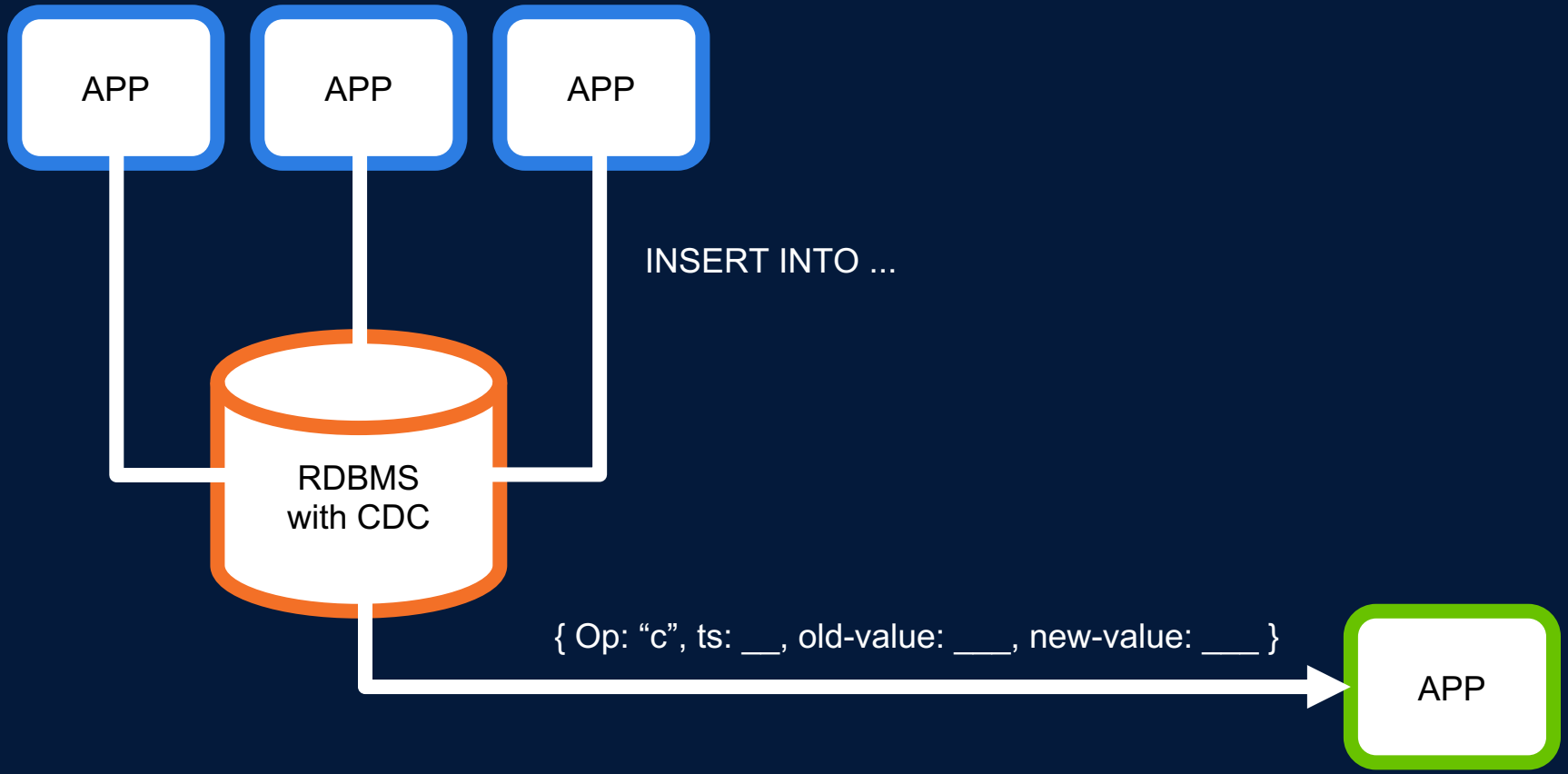
Push over Pull

Easier to integrate new systems

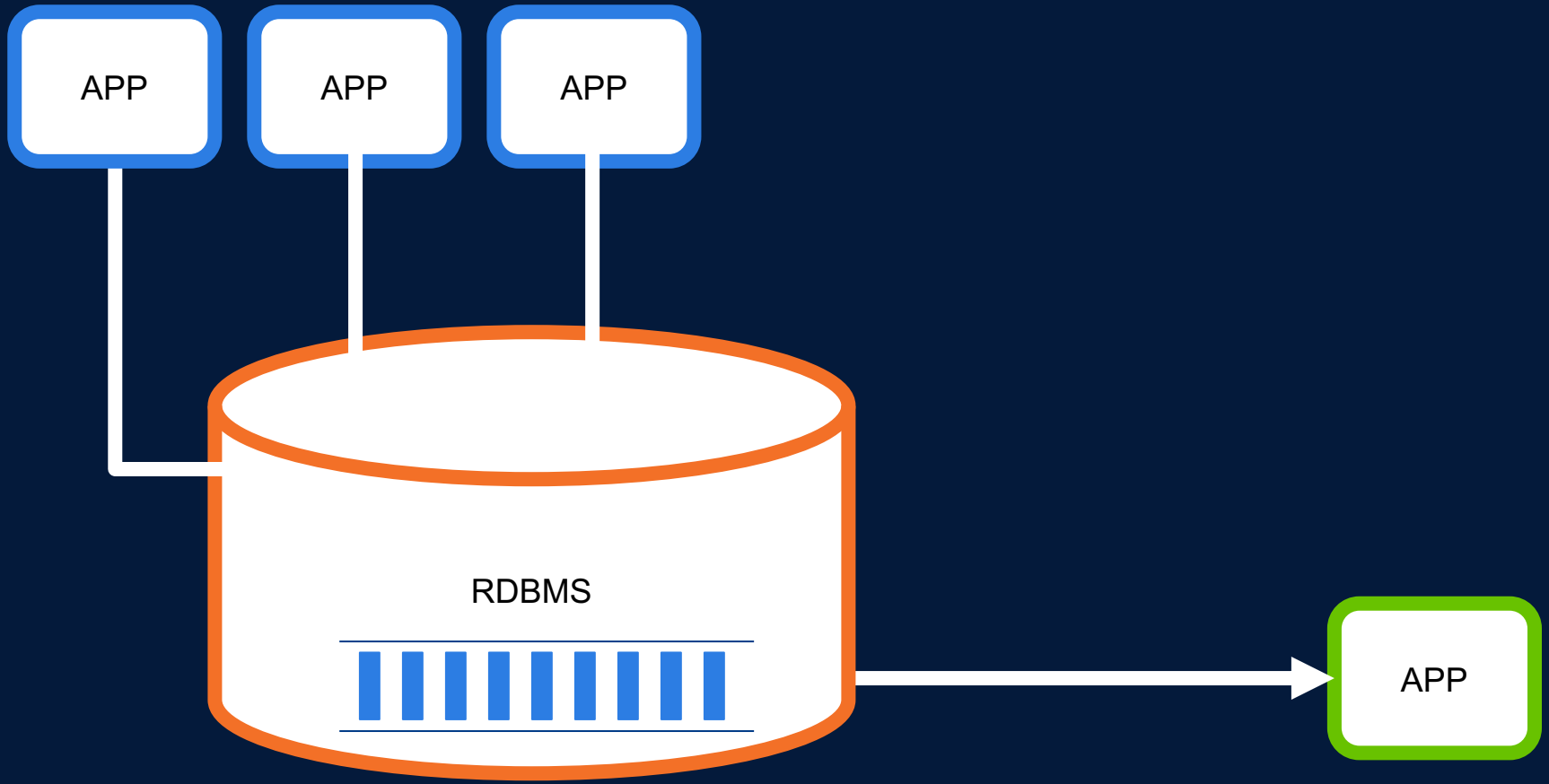
OH, YEAAHH!

IMPEDANCE MISMATCH

Change Data Capture (CDC) turns legacy database to a source of event stream.


APP
APP
APP
INSERT INTO ...
RDBMS
with CDC
{ Op: “c”, ts: __, old-value: ___, new-value: ___ }
APP

hazelcast

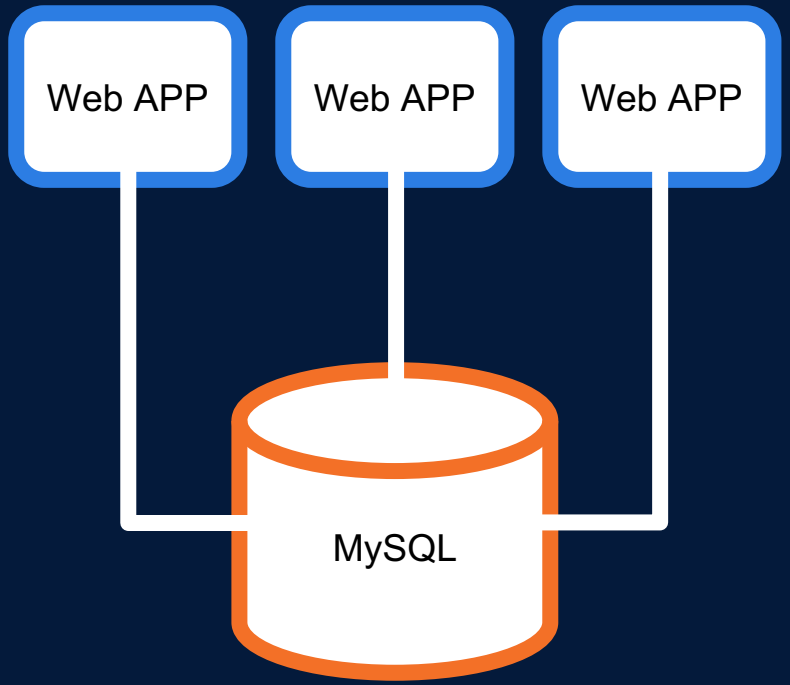


hazelcast

Travelling Back in Time

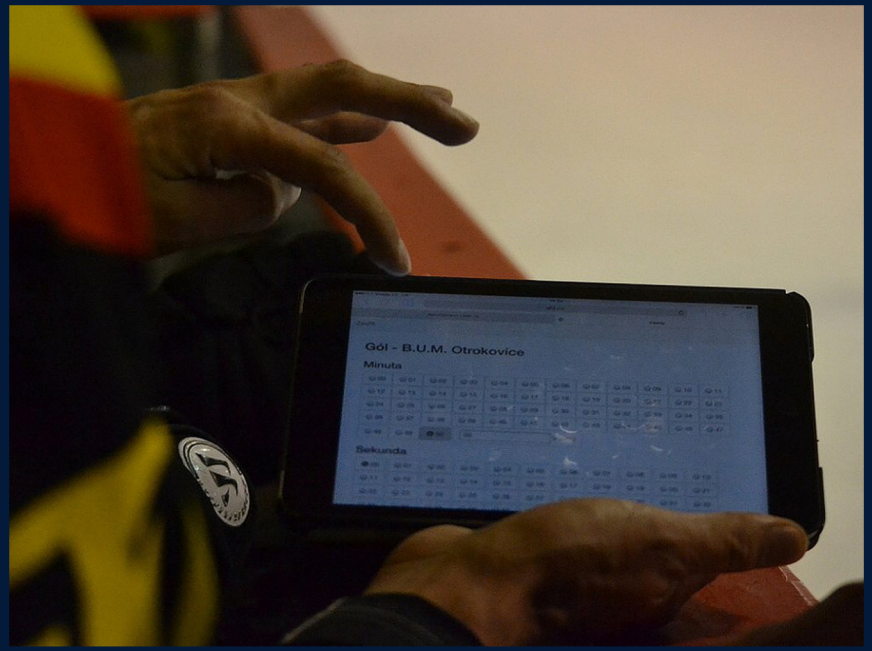
https://github.com/hazelcast/hazelcast-jet-demos/tree/master/debezium-cdc-without-kafka

hazelcast



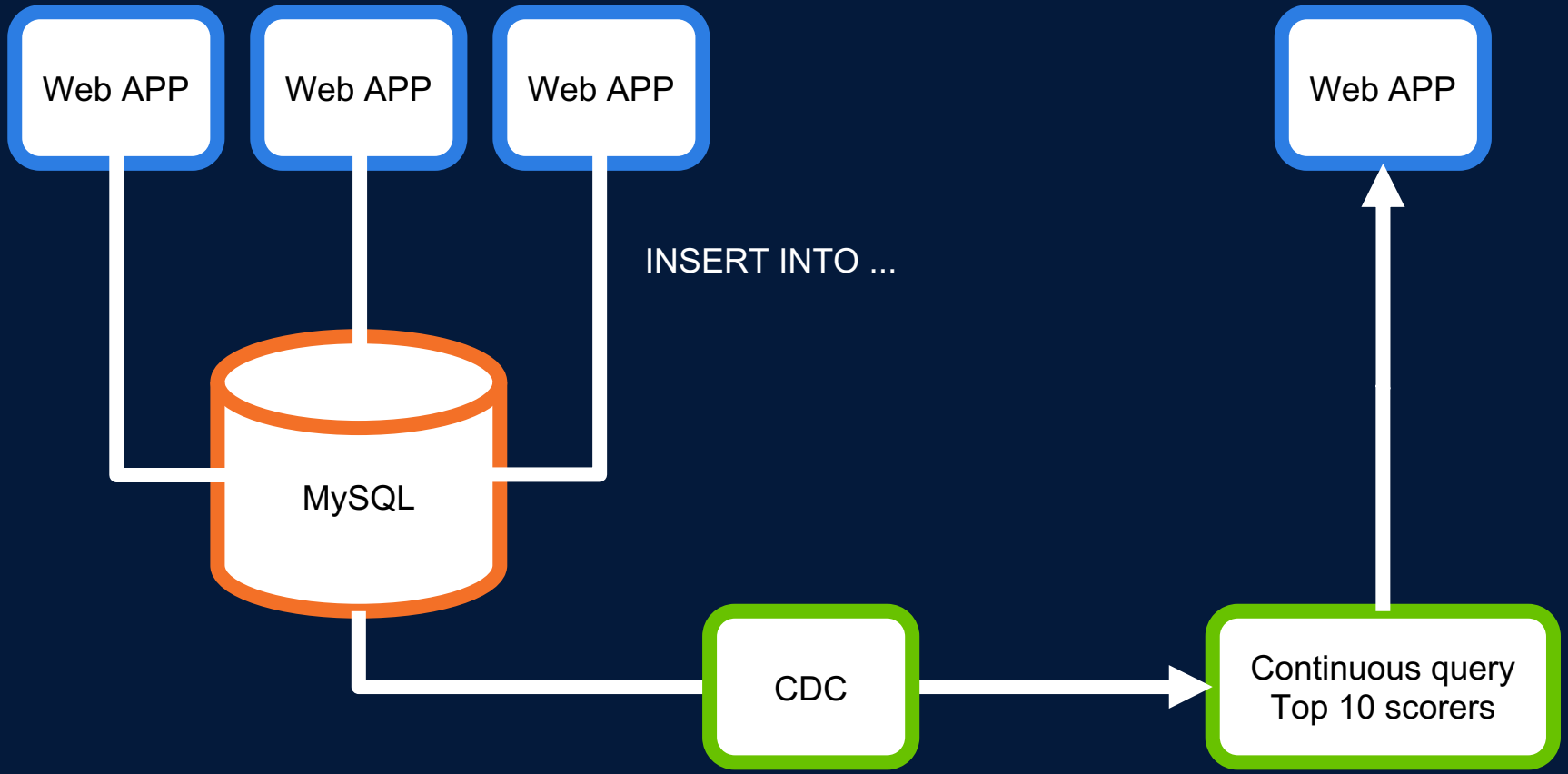
LAMP stack
GBs of data
Mostly OLTP, OLAP after hours

hazelcast



Real-time updates = Real-time expectations

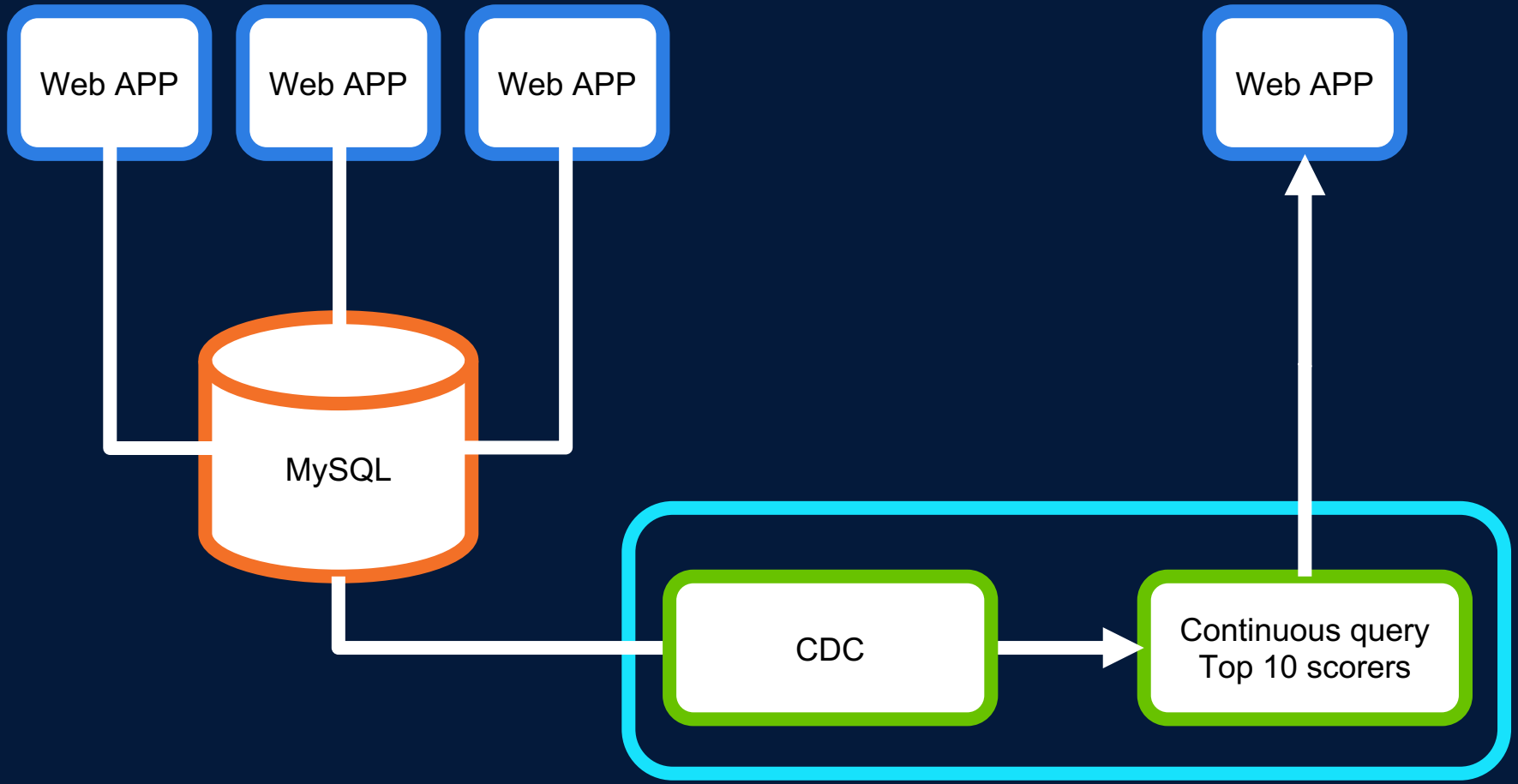
hazelcast


Web APP
Web APP
Web APP
Web APP
INSERT INTO ...
MySQL
CDC
Continuous query
Top 10 scorers

hazelcast

Open-source library with stream processor, connectors including CDC and a in-memory key-value store.

Single Java Binary
Elastic Clustering
No ZooKeeper, HDFS..
Just Java 8 and above

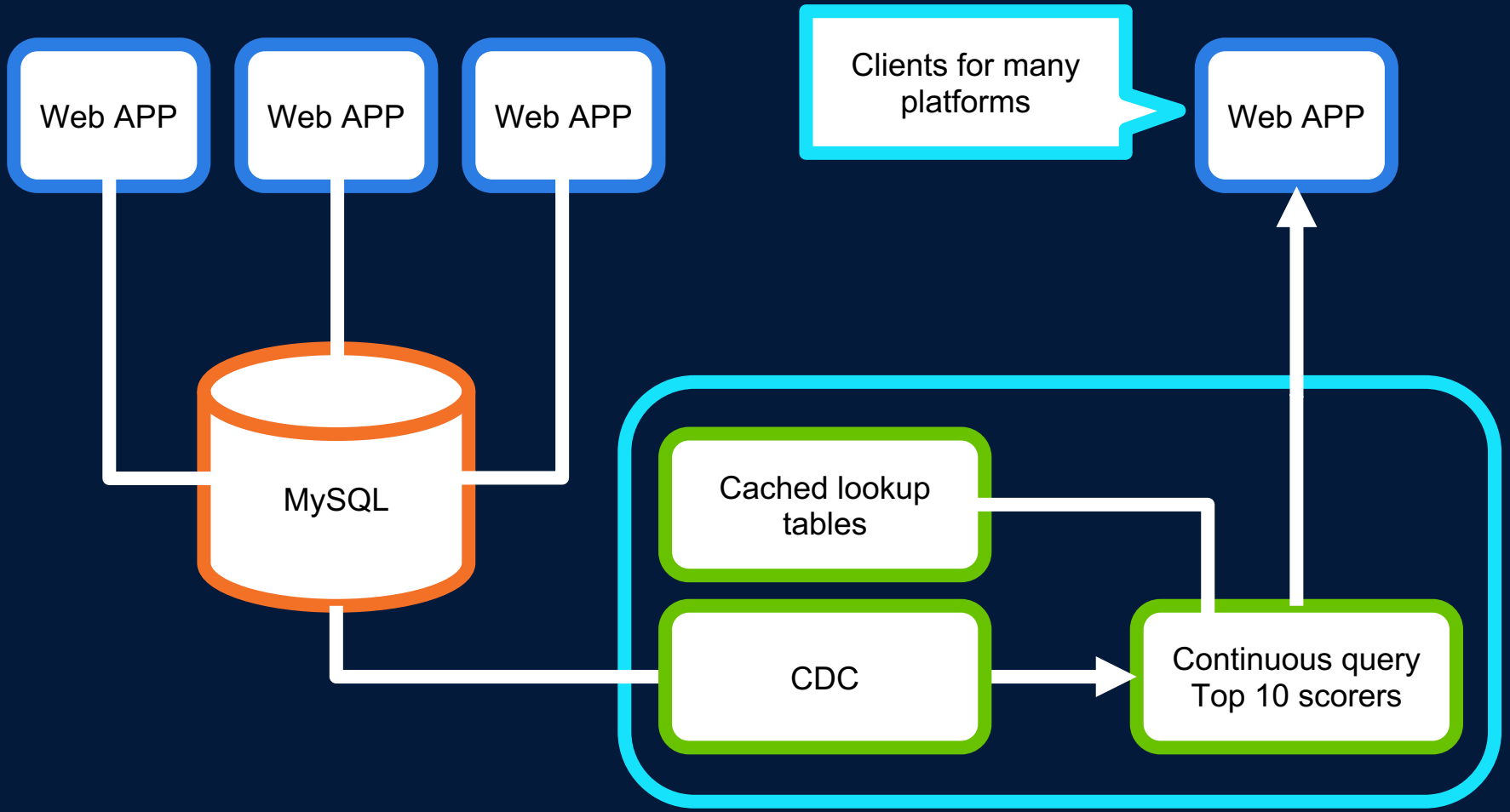

Web APP
Web APP
Web APP
Web APP
MySQL
CDC
Continuous query
Top 10 scorers

hazelcast

Hazelcast Jet runs Debezium for CDC

```
# The following can be used as easy to replay
# backup logs or for replication.
server-id               = 1
log_bin                 = /var/log/mysql/mysql-bin.log
binlog_format      = row
binlog_row_image   = full
expire_logs_days        = 2
```

Relational databases usually keep shorter history, compared to dedicated log-based storages.


Web APP
Web APP
Web APP
Clients for many platforms
Web APP
MySQL
Cached lookup tables
CDC
Continuous query Top 10 scorers

hazelcast

Event-driven end2end

hazelcast

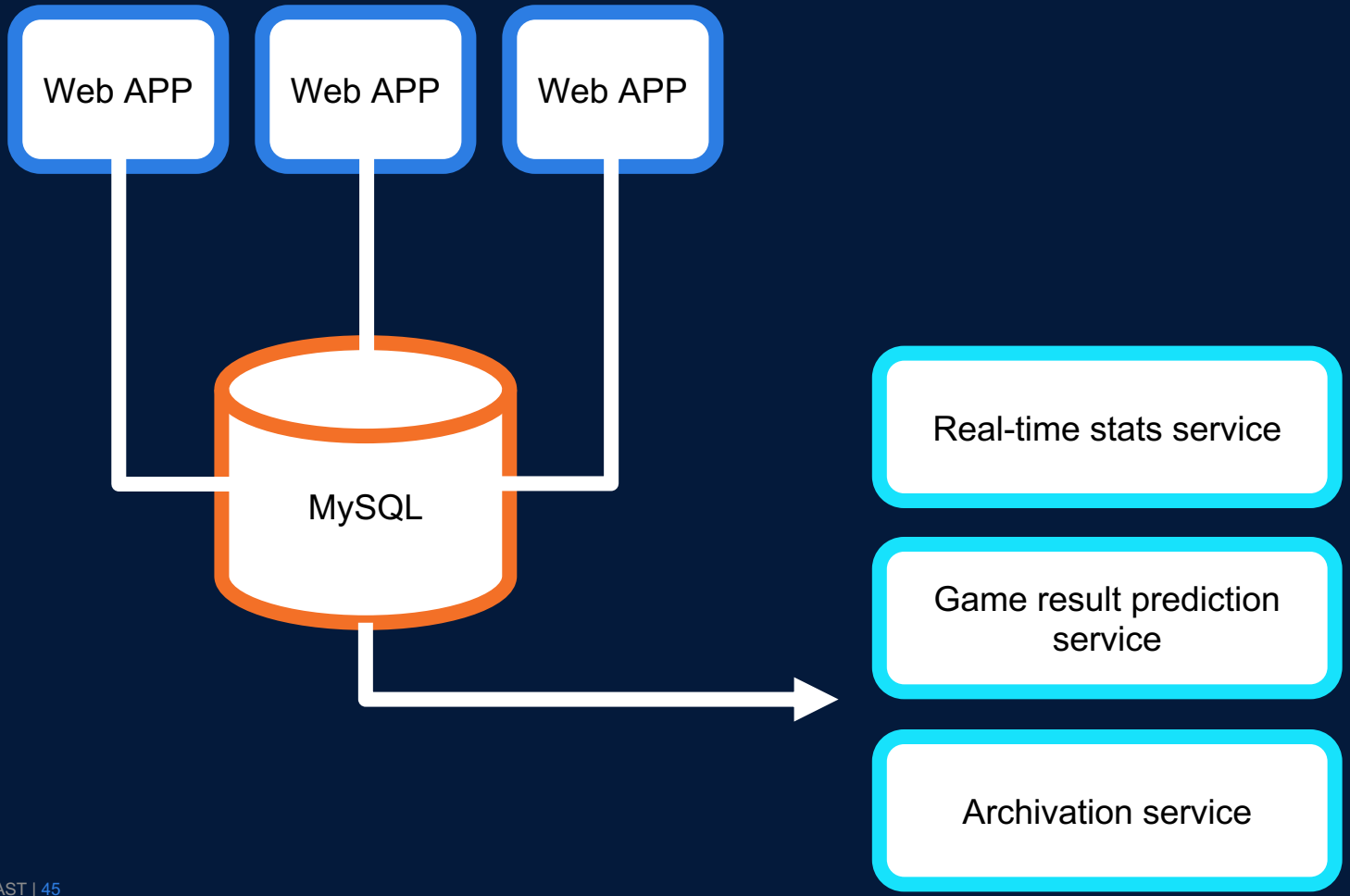
CDC + SPE + CACHE =

Materialized Views Offloaded from the DB

hazelcast

Modularization Microservices

hazelcast


Web APP
Web APP
Web APP
MySQL
Real-time stats service
Game result prediction service
Archivation service

hazelcast

Scaling

hazelcast

Conclusions

Extract change stream from your legacy stack using CDC.

Use CDC and stream processing to create materialized views outside the database to reduce the database load and modularize your architecture.

Hazelcast Jet is an open-source Java library to do the job.

- CDC
- Streaming
- Caching

Questions

Visit our stand on 3rd floor.

jet.hazelcast.org