

AEROSPIKE  
SUMMIT '19

Quantcast

# Rebuilding a Realtime Feature Store

Kristi Tsukida  
Staff Engineer  
Quantcast







# Quantcast

(2006) Internet analytics



(2009) Online advertising

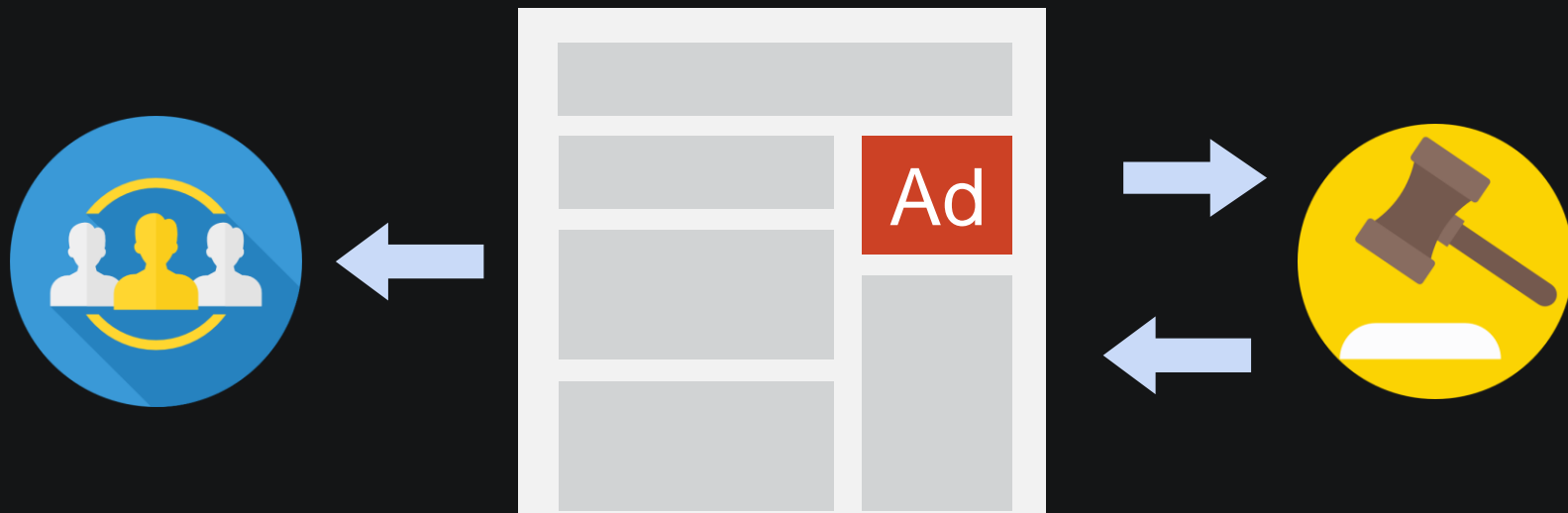


(2012) Real-time scoring

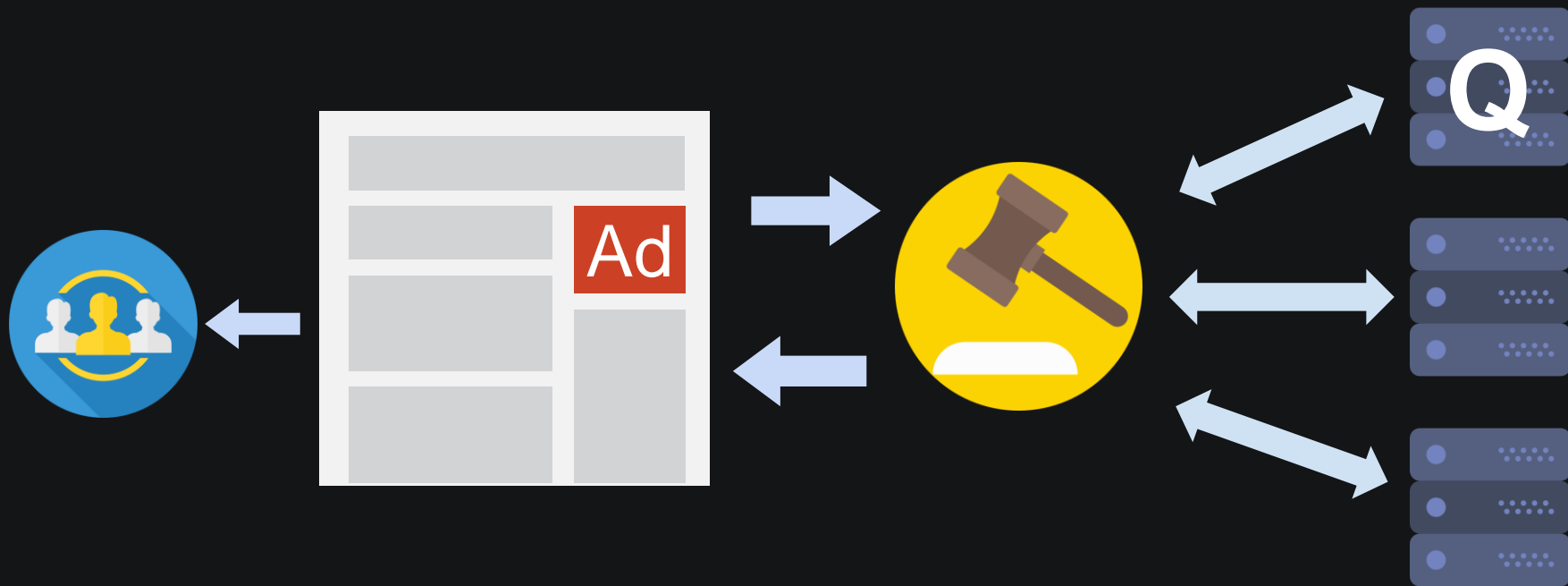
(future) **The Definitive  
Online Audience Platform**



# Advertising & Real Time Bidding (RTB)

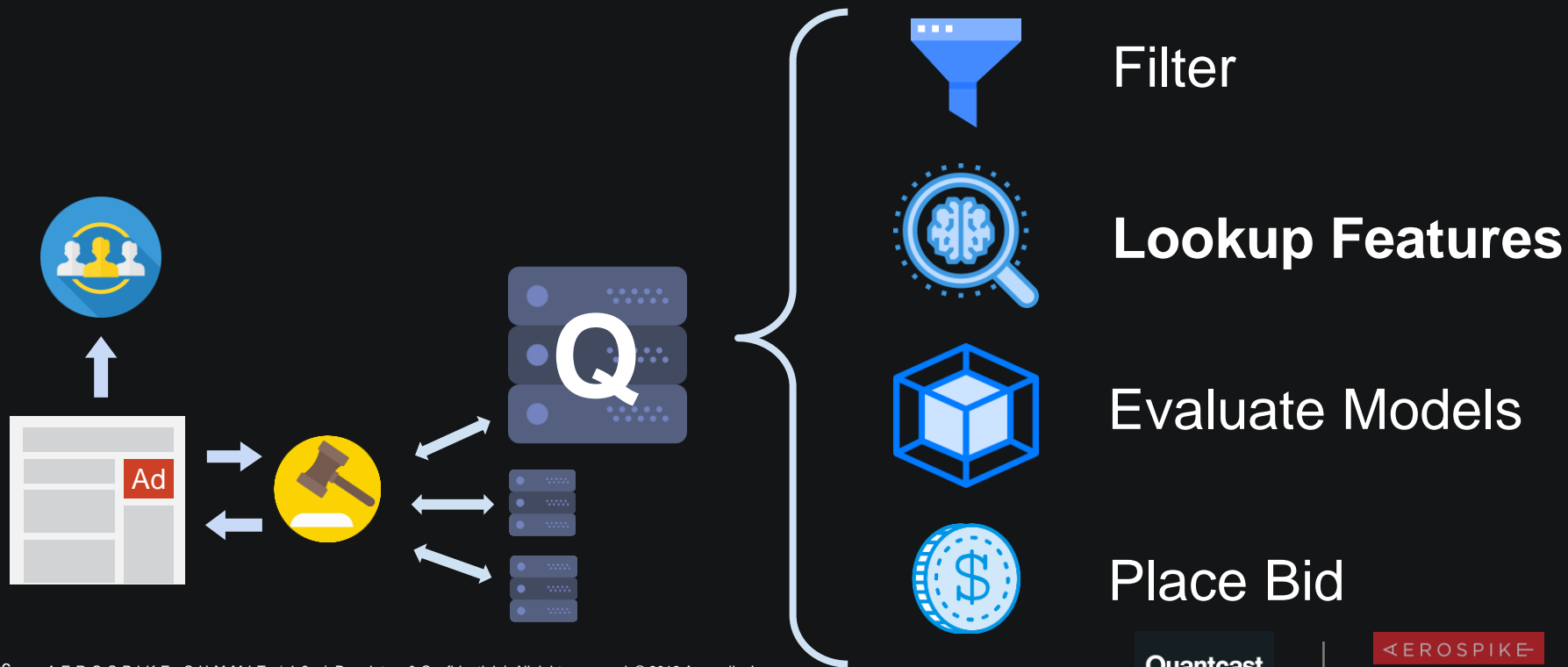


# Advertising & Real Time Bidding (RTB)

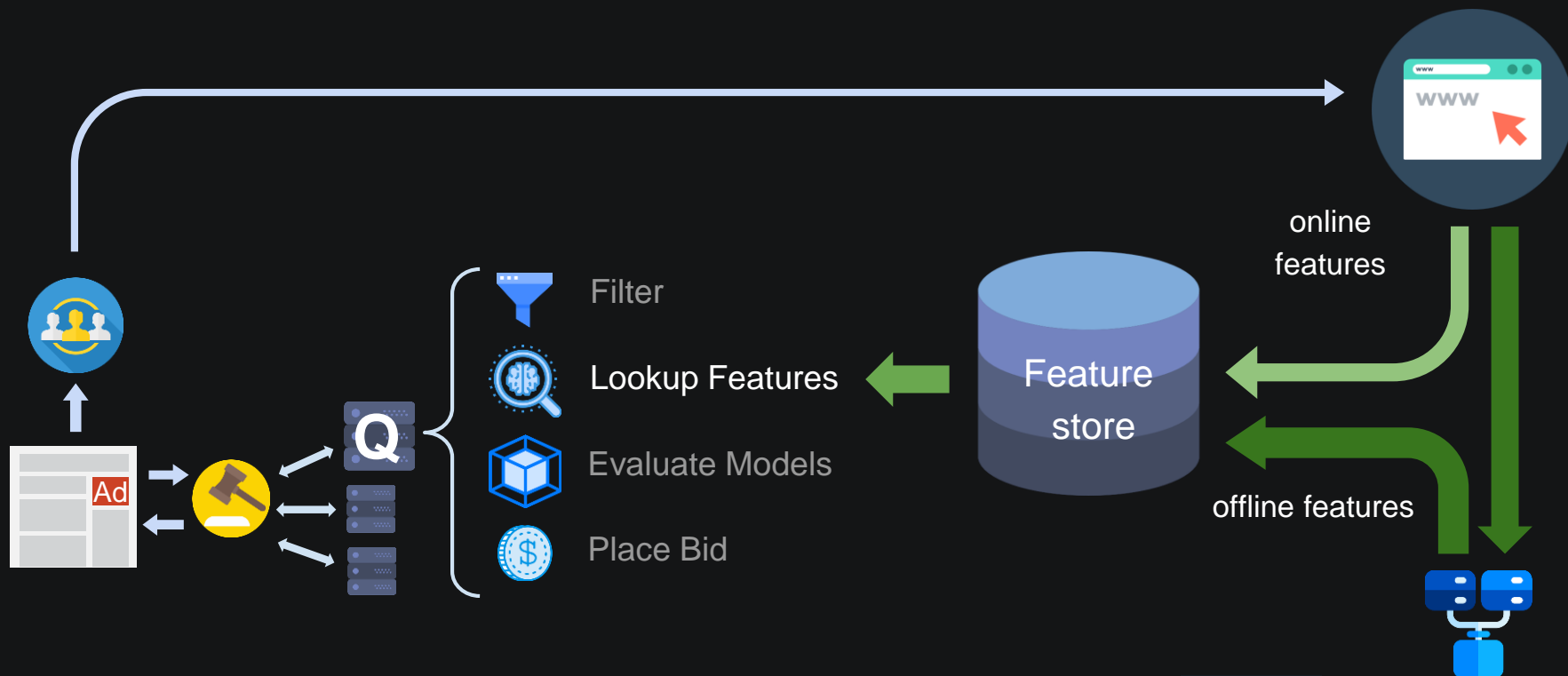




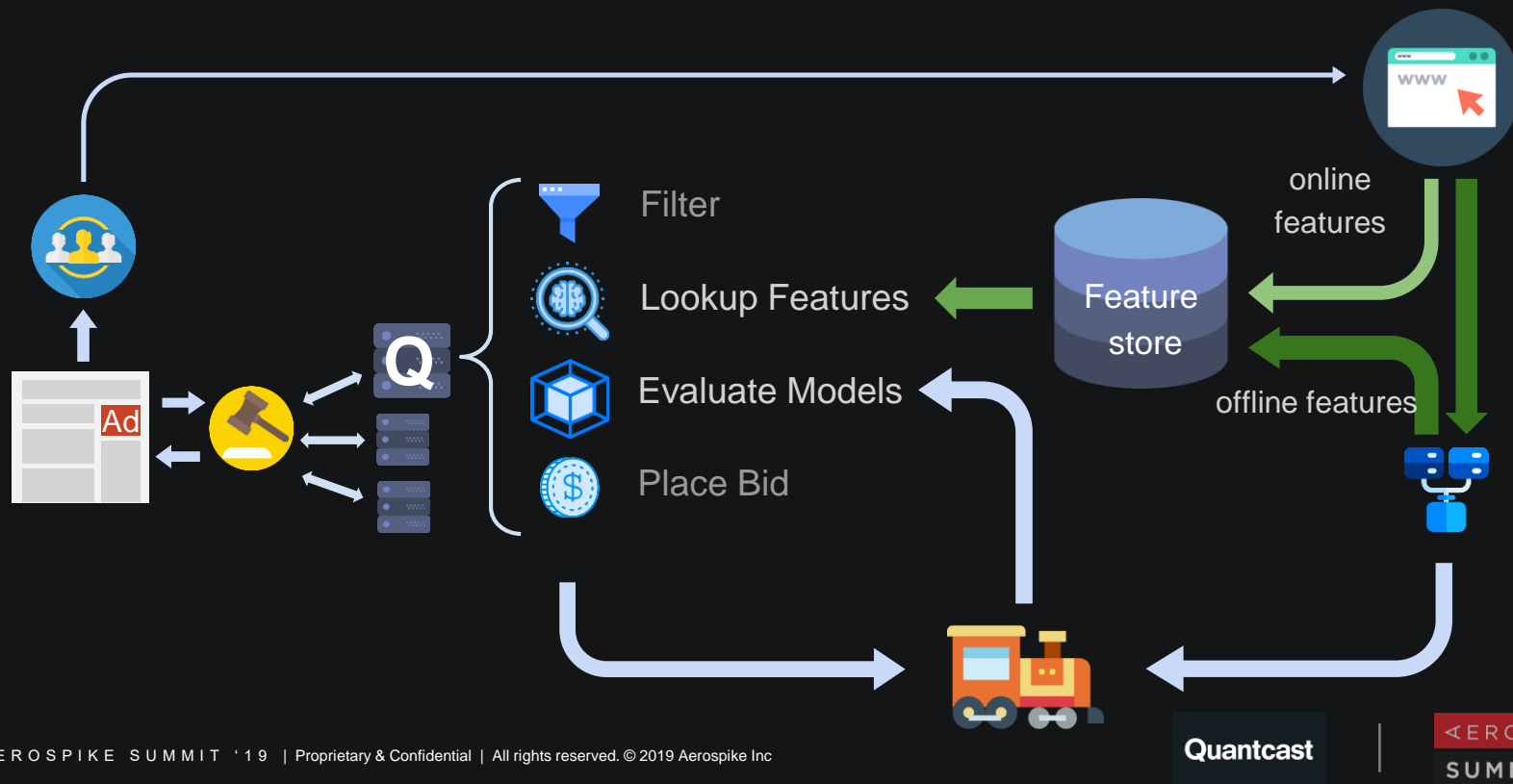
# Advertising & Real Time Bidding (RTB)



# Advertising & Real Time Bidding (RTB)

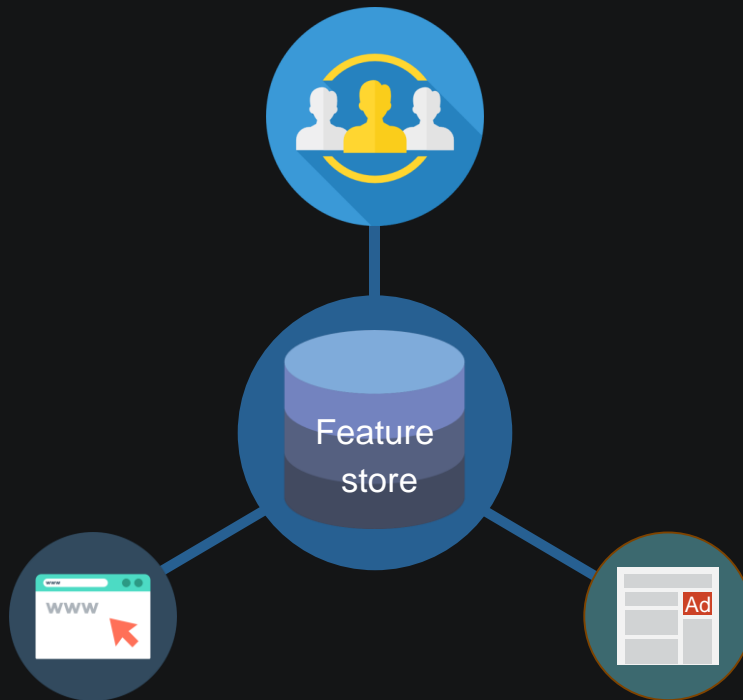


# Advertising & Real Time Bidding (RTB)





# Realtime Feature Store



# Legacy Feature Store

- not flexible
- scaling and operational issues
- unreliable / inconsistent data
- loss of institutional knowledge
- no clean service APIs
- stores other data sets



# New Feature Store

- enable feature experimentation
- accommodate new audience platform
- improved reliability
- be prototype for other bidding data stores



# Rebuild vs Refactor

## Business context

- Velocity
- Technology
- Maintainability



# Build vs Buy



- Specialization value
- Business alignment
- Maintainability
- Cost
- People investment

# AEROSPIKE

- Low latency
- Optimized SSD for large data size
- Scalable key value store

# Proof of Concept

Does this meet our needs?

Is it worth investing in?

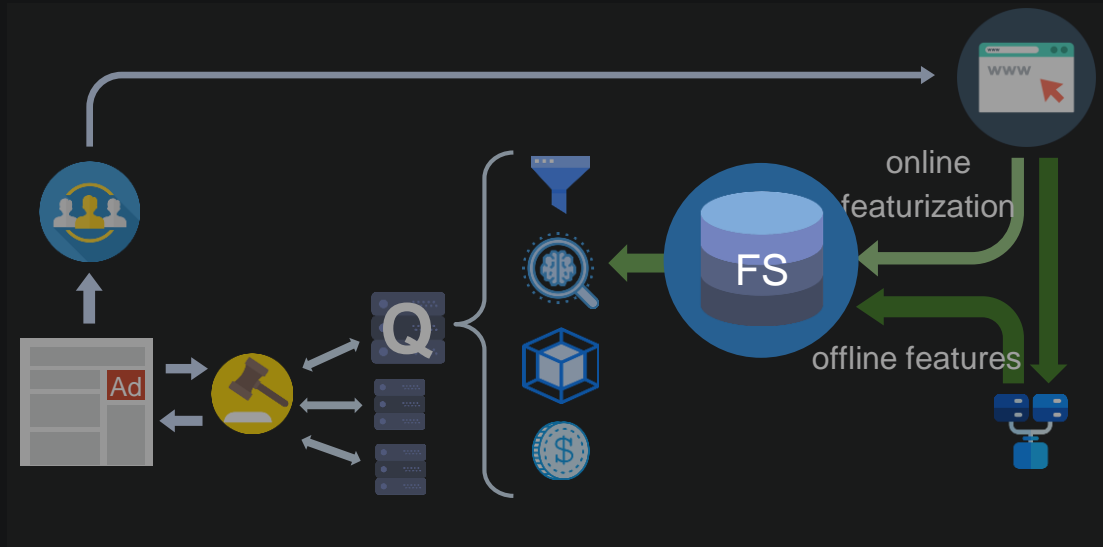
- limited scope
- take (calculated) short cuts
- get results





# Feature Store Requirements

- 10 Billion records
- 8Tb data
- 2ms lookups
- 1M lookups/second
- 200k updates/second



# Data Layout

- Serialized byte blobs (e.g. protobufs)
  - must perform read for all updates
- Aerospike data objects (bins, maps, lists)
  - can do some updates in a single op (e.g. expiring features)

# Aerospike Proof of Concept Evaluation

- Latency
  - p99: 5ms (legacy: 25ms)
  - p95: 2ms (legacy: 10ms)
  - p50: 1ms (legacy: 1ms)
- Reliable
  - 100% of writes successful
- Easy to setup
- Good documentation
- Excellent support

Thanks!



