



# Citrusleaf

Making Web Scale  
Easy and Affordable



# Display advertising

- Lie:
  - **Content** determines the value of an impression
- Lie:
  - The value of an individual user is constant
  - The decision **funnel**
- Truth:
  - The **right ad**, to the **right user**, at the **right time**



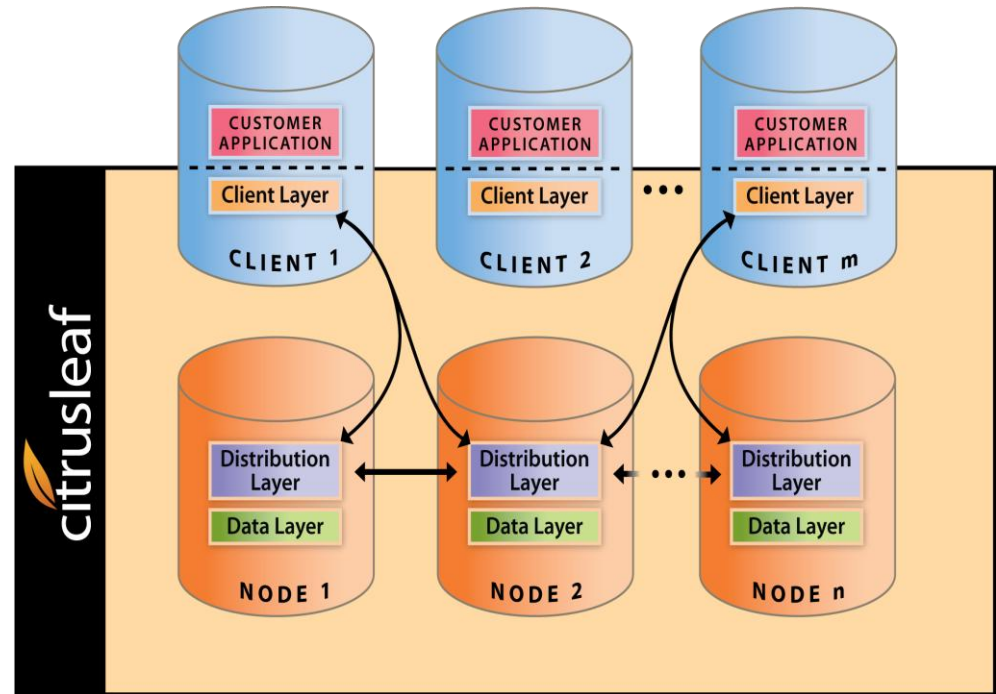
# Today's state of the art

- Impression logs => user segments (warehouse)
- Third part data => user segments
- Region segments
- Real-time per-user factor
  - Frequency caps
  - Session management for cookie-less users
- Mapping external partner IDs
- Build optimization tables based on log analysis

**Simple math determines highest value ad**

# Real-time data requirements

- Scalable and flexible
- 100% availability
- Billion object support
- High performance with low hardware cost
- Sophisticated data eviction policies





# Case study

## Map store

Key      cookie string or  
          partner id  
Value    internal user id (8 bytes)

### North America

1B ~ 1.5B objects  
Write load:  
    1k ~ 2k per second  
Read load:  
    10k ~ 50k per second

### Configuration

DRAM backed by disk  
Lowest latency (0.4 ms)

## User store

Key      internal user id  
Value    Segment data  
          Frequency caps  
          Other optimization data

### North America

500M ~ 800M objects  
Write load:  
    10K ~ 20K per second  
Read load:  
    20k ~ 100k per second

### Configuration

1T ~ 4T user storage  
Low latency (0.8 ms)



# Market differentiation

- Control over impressions
- Quality and variety of inventory
- Better insight generation & tracking
- And:
  - High quality user understanding

**Trend: multiple user stores**



# The next challenge

Weakness: “last ad takes all”

Performance is judged externally by the last advertiser before a conversion

So... game the system to place strategically

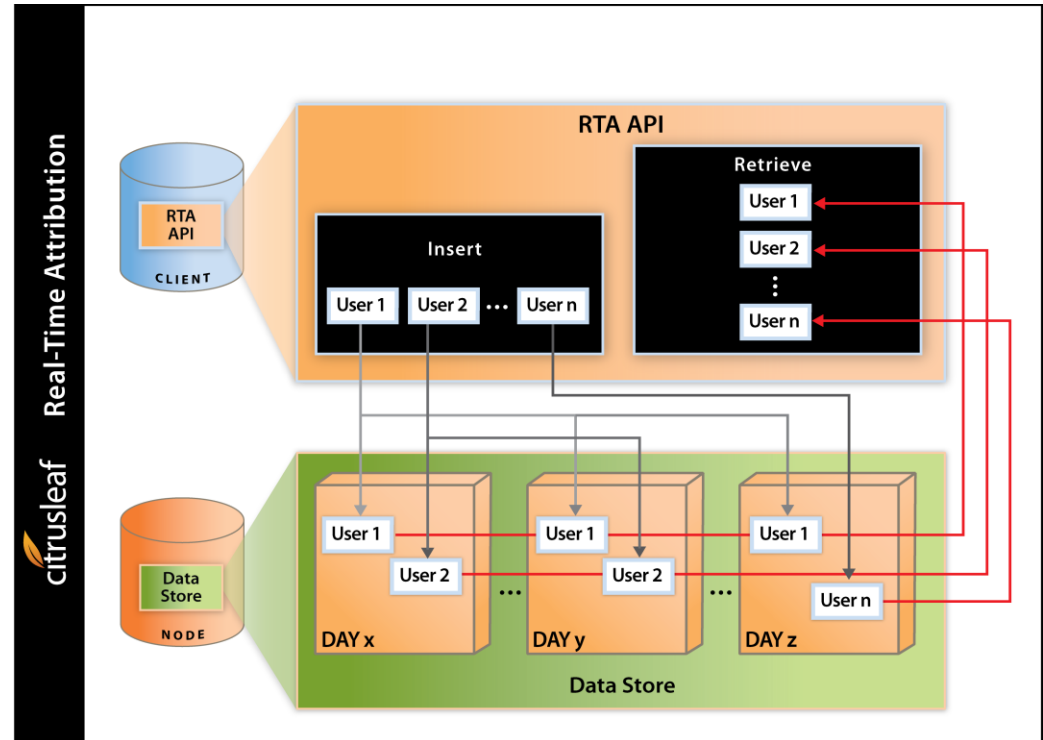
New technology allows real-time understanding of user behavior

But ...

Will advertisers stand for it?

# Indexed user behavior storage

- Store 100B+ user behavior objects (1B per day, 90 days)
- Index by user\_id, giving sub-second access to entire behavior chain
- Optimized for rotational media
- Time-based eviction





# Indexed user behavior storage

- Solves the “needle in a haystack” problem
- Allows immediate behavioral triggering
- Removes ETL and Map/Reduce scans
- Allows sophisticated attribution models,  
(and advanced optimization and reporting)

**The right ad,  
to the right user,  
at the right time**



# Citrusleaf

Making Web Scale Easy and Affordable

Contact Info: [brian@citrusleaf.net](mailto:brian@citrusleaf.net)