MOLOCO

Machine Learning and data infrastructure in Ad-Tech by Moloco

How to handle 80B+ real-time bidding everyday

Speaker: Min-gyu Cho

Work Experience

- Data Scientist / Head of Ad Cloud Solution at Moloco Inc.
- Assistant Professor at DGIST
- Data Science Unit Leader at 11st
- Product Manager & Big Data Strategy at SK Planet
- Associate at McKinsey & Company

Education

- M.S. & Ph.D. in Computer Science and Engineering at Univ. of Michigan
- B.S. in Electrical Engineering at Seoul National Univ.

Speaker: Daeseob Lim

Work Experience

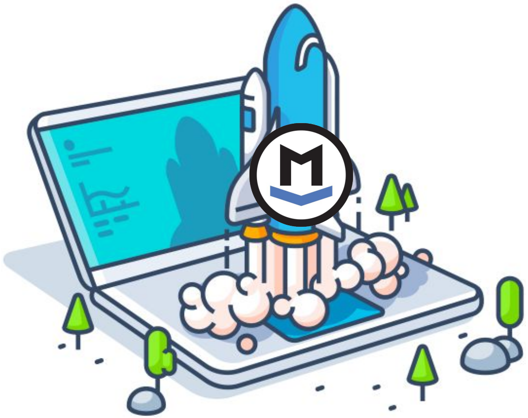
- Technical Program Manager at Moloco
- ML software engineering for VOD streaming & IPTV solutions at Castis.

Education

- Ph.D. candidate in Brain & Cognitive Sciences at Seoul National Univ.
- M.S in Computer Science at Univ. of California, San Diego
- B.S. in Computer Engineering at Seoul National Univ.

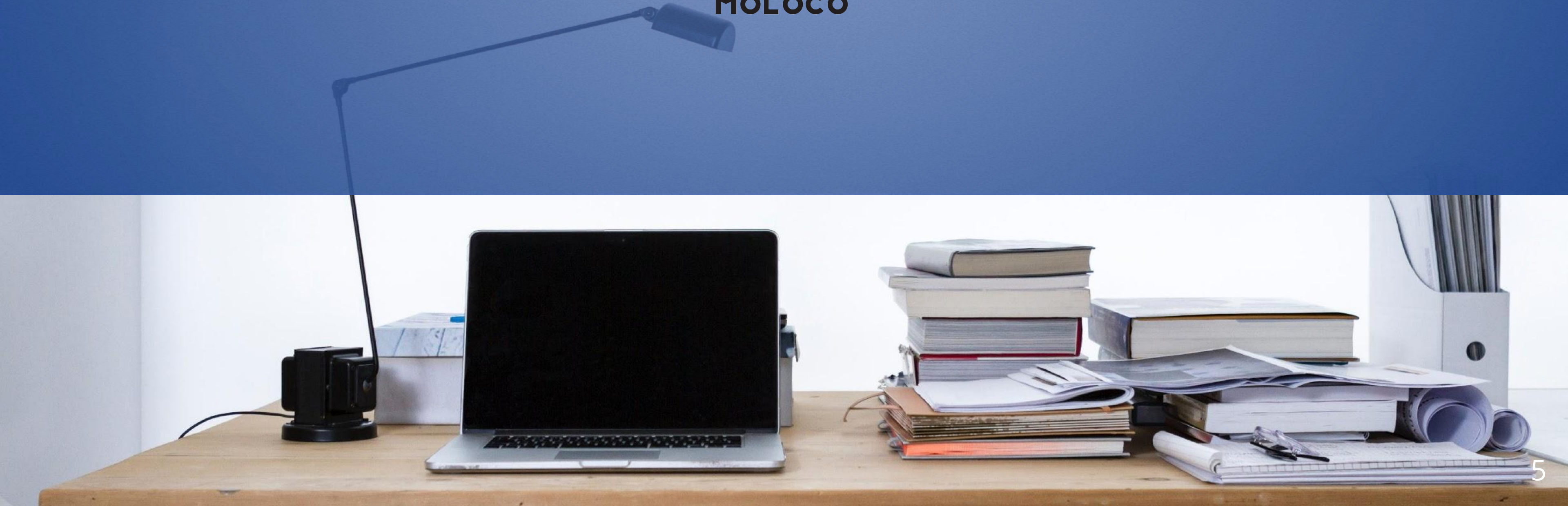
Today’s Agenda

MOLOCO



Ad-tech 101

MOLOCO



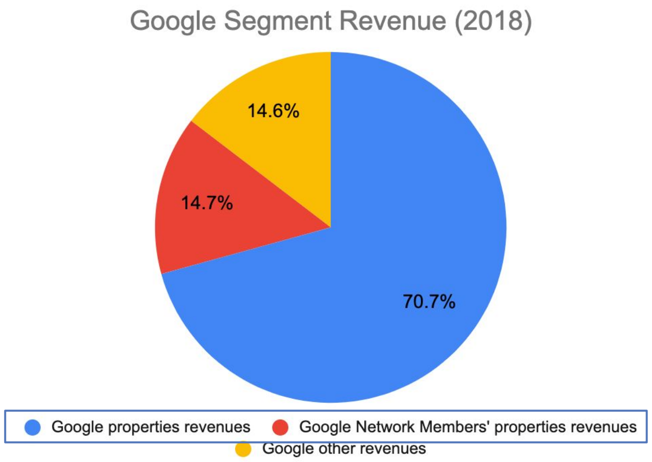
Ad-Tech?

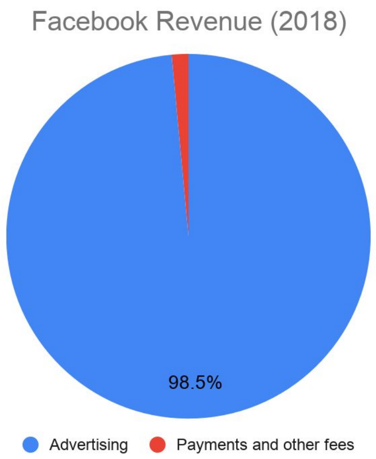
Advertising + Technology

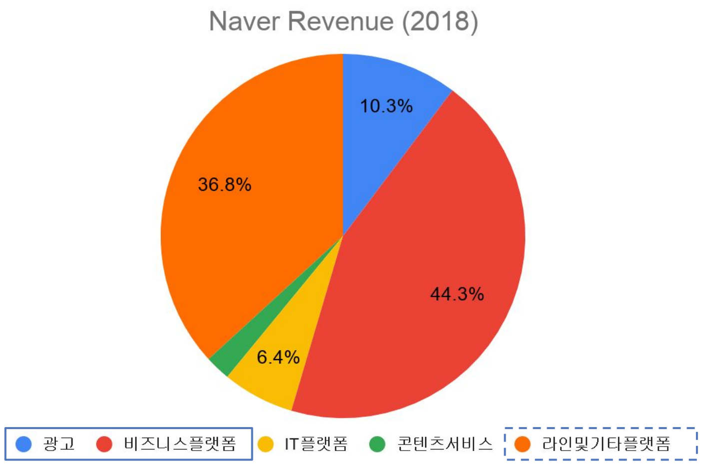
Why do I need to pay attention to ad-tech?

Advertising is the main revenue source of tech giants

Advertising Revenue







What is advertising?

“**Advertising** is an audio or visual form of *marketing communication* that employs an openly sponsored, nonpersonal message to *promote* or sell a *product, service or idea*.” - Wikipedia

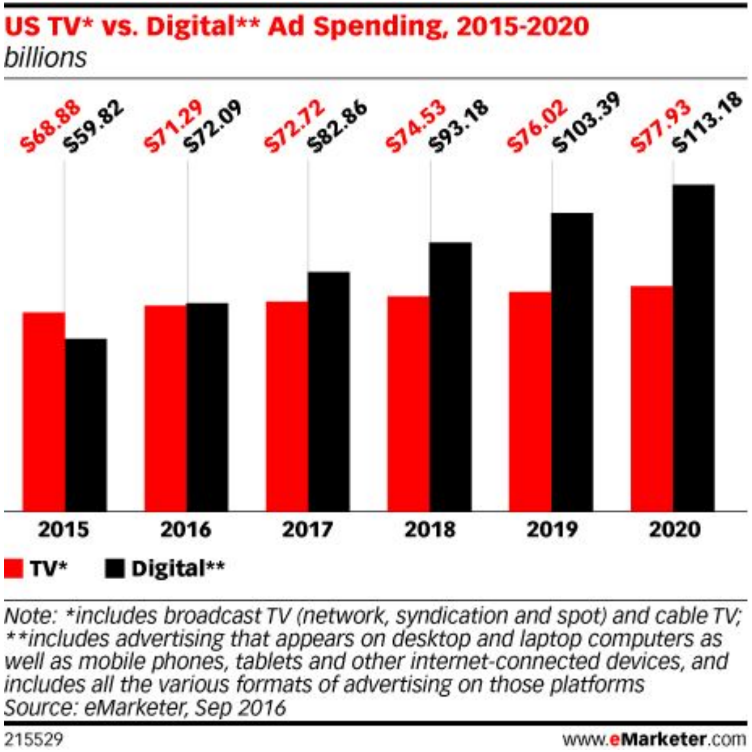
Needle shop ad
From Song dynasty,
China (1,000+ years ago)

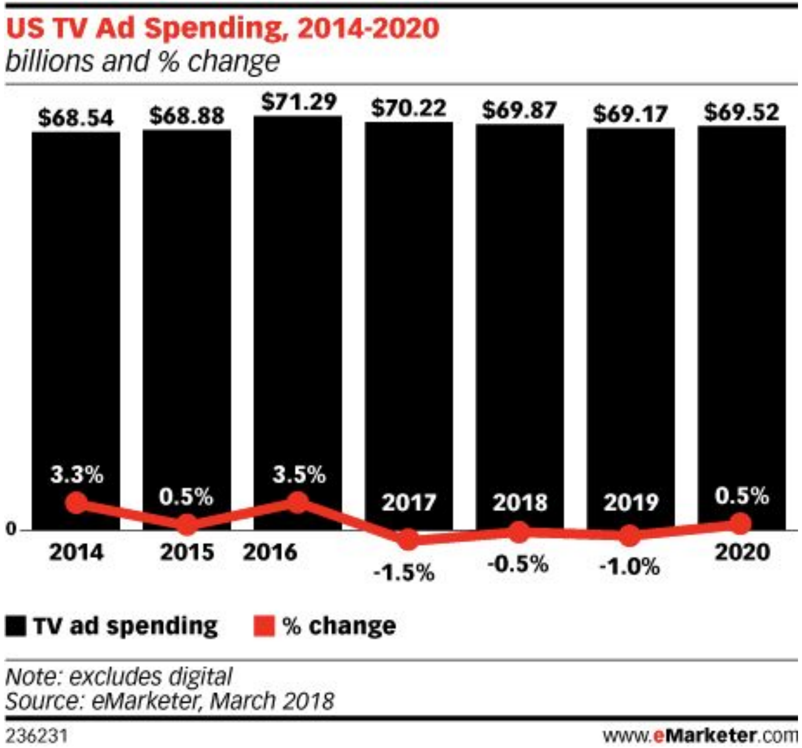
Coca-cola ad
from 1890

TV ad

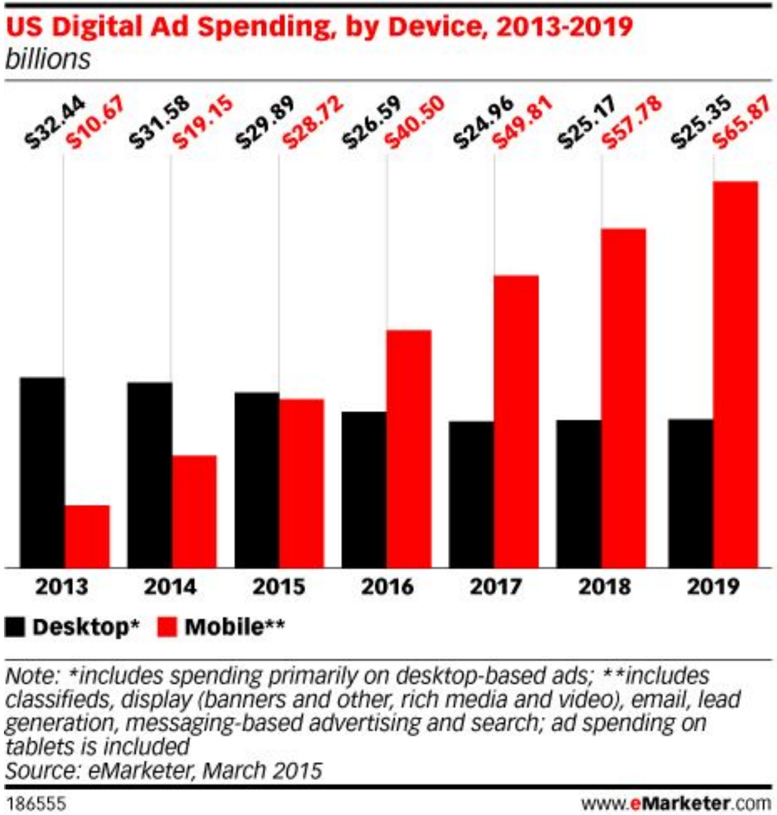
Digital ad

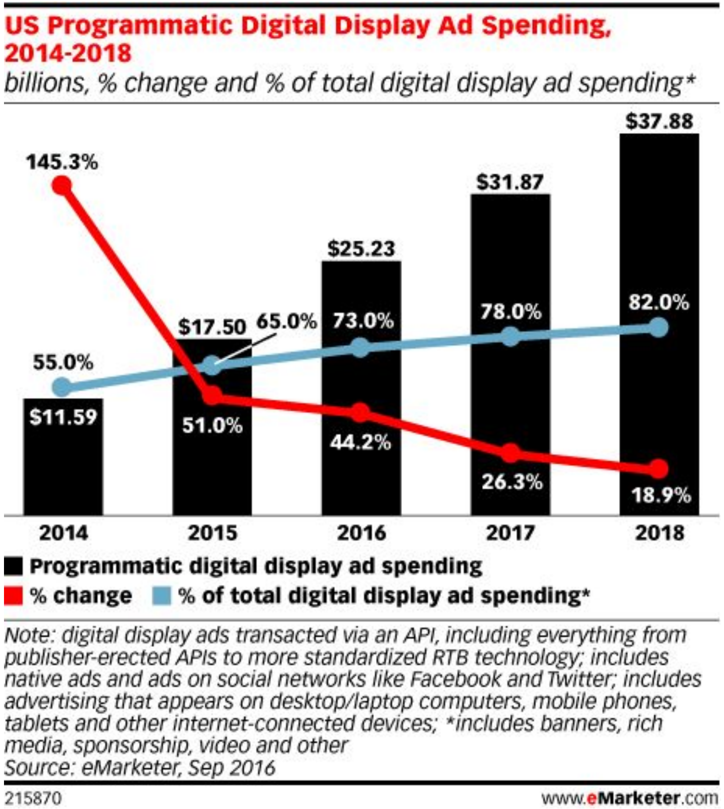
Digital is eating the advertising market





Mobile & programmatic is eating the digital ads market





What is programmatic ads?

What is programmatic ad buying?

… the use of *software* to purchase digital advertising, …. It's using machines to buy ads, basically.

Why does programmatic advertising matter?

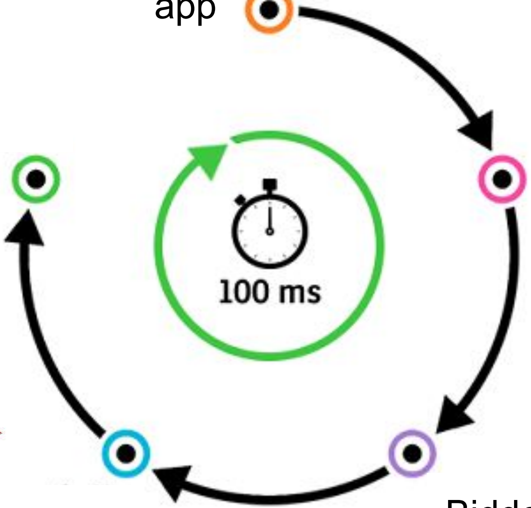
Efficiency. … … Programmatic advertising technology promises to make the ad buying system more efficient, and therefore cheaper, by removing humans from the process wherever possible. …

Source: https://digiday.com/media/what-is-programmatic-advertising/

So what problem do we have?

bidrequest is born when user opens an app

Winning ad
is served
at user's device



Exchange broadcasts "***bidrequest***" to potential bidders

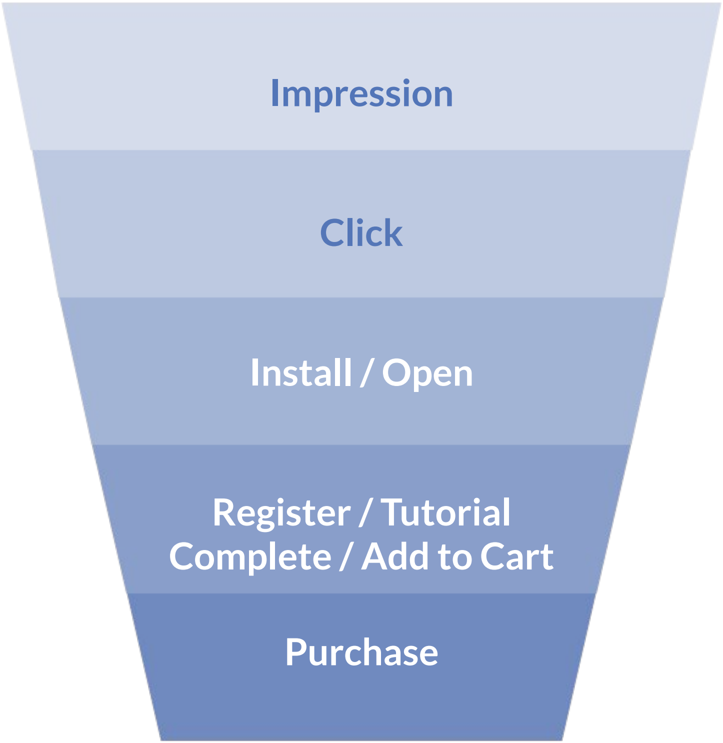
Auction is over
at exchange

Bidders make bids

What is the goal of the advertiser?

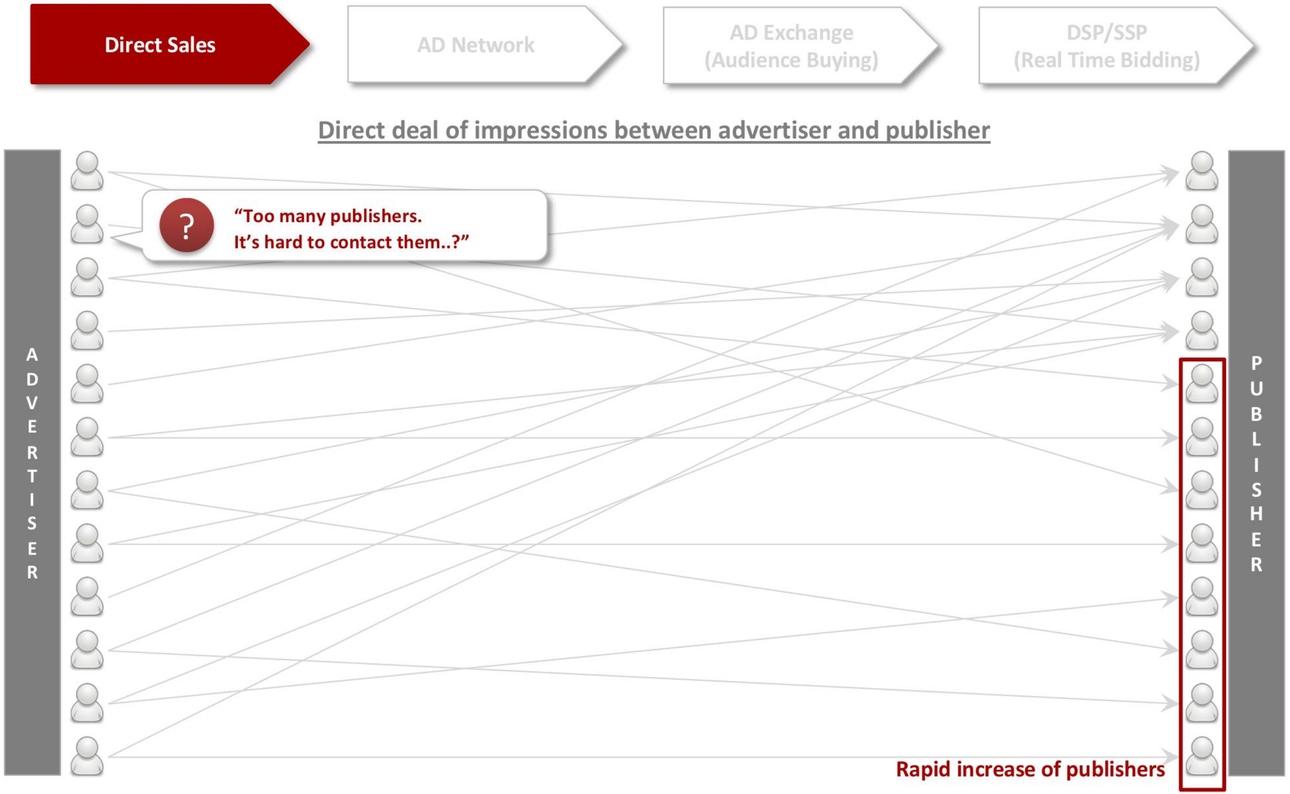
Advertisers want user to reach further in the marketing funnel, e.g.,

- Awareness of their applications/services
- Response to advertising
- Install
- Retention
- Purchase

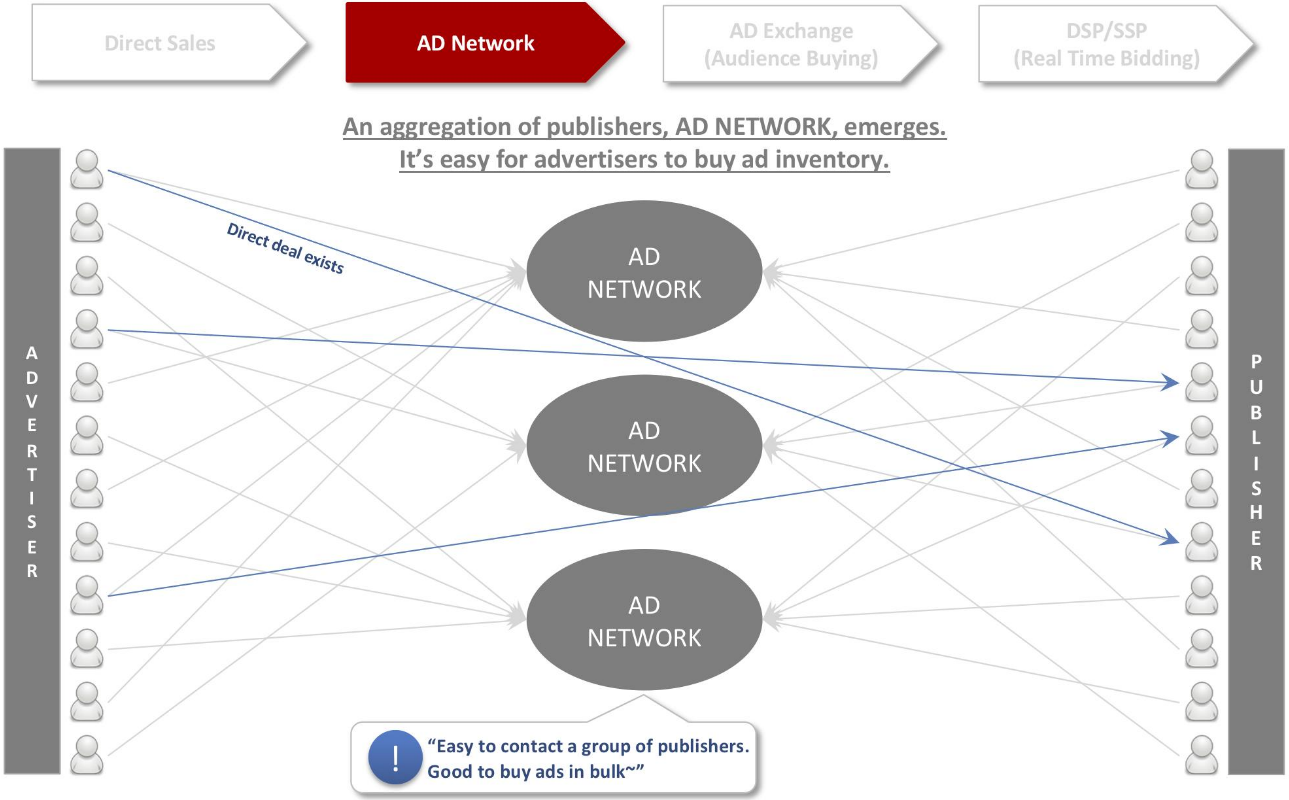


Evolution of Digital Ad Platform

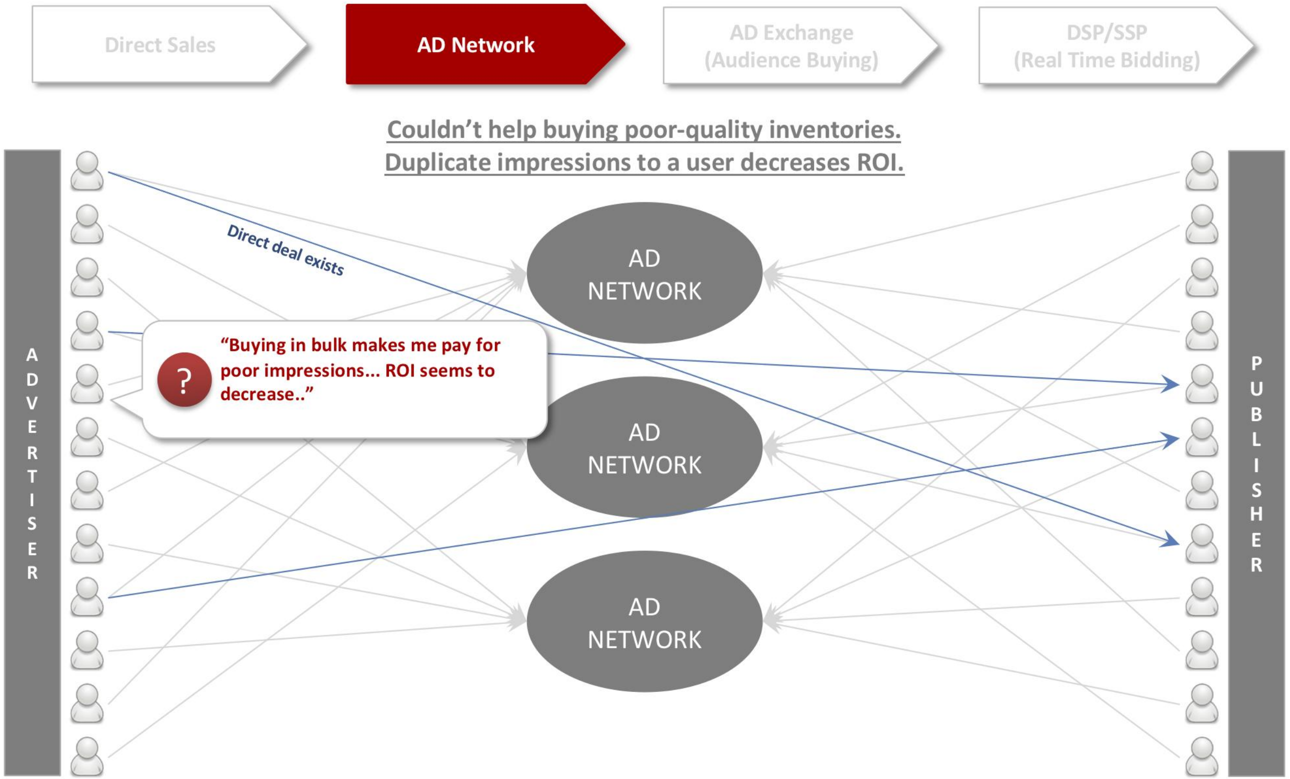
Evolution of digital ad platform: direct sales



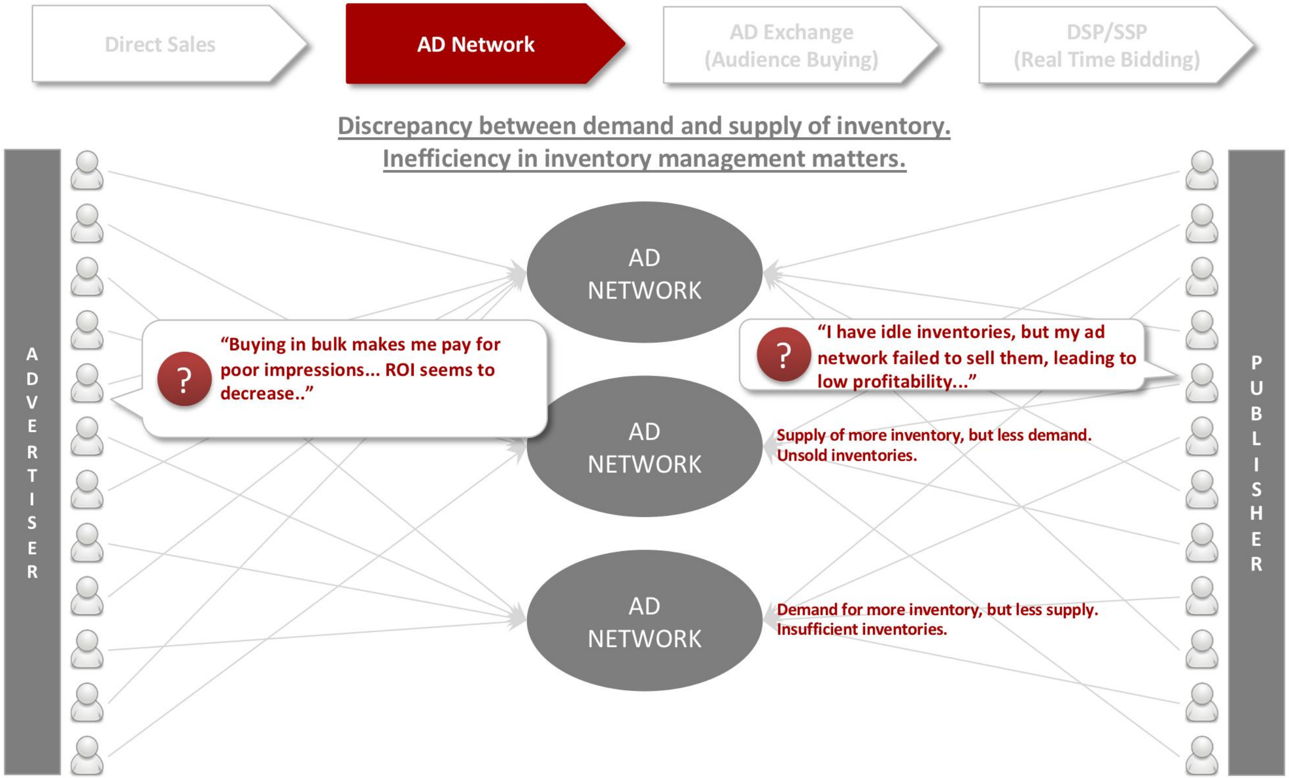
Evolution of digital ad platform: ad network (1/3)



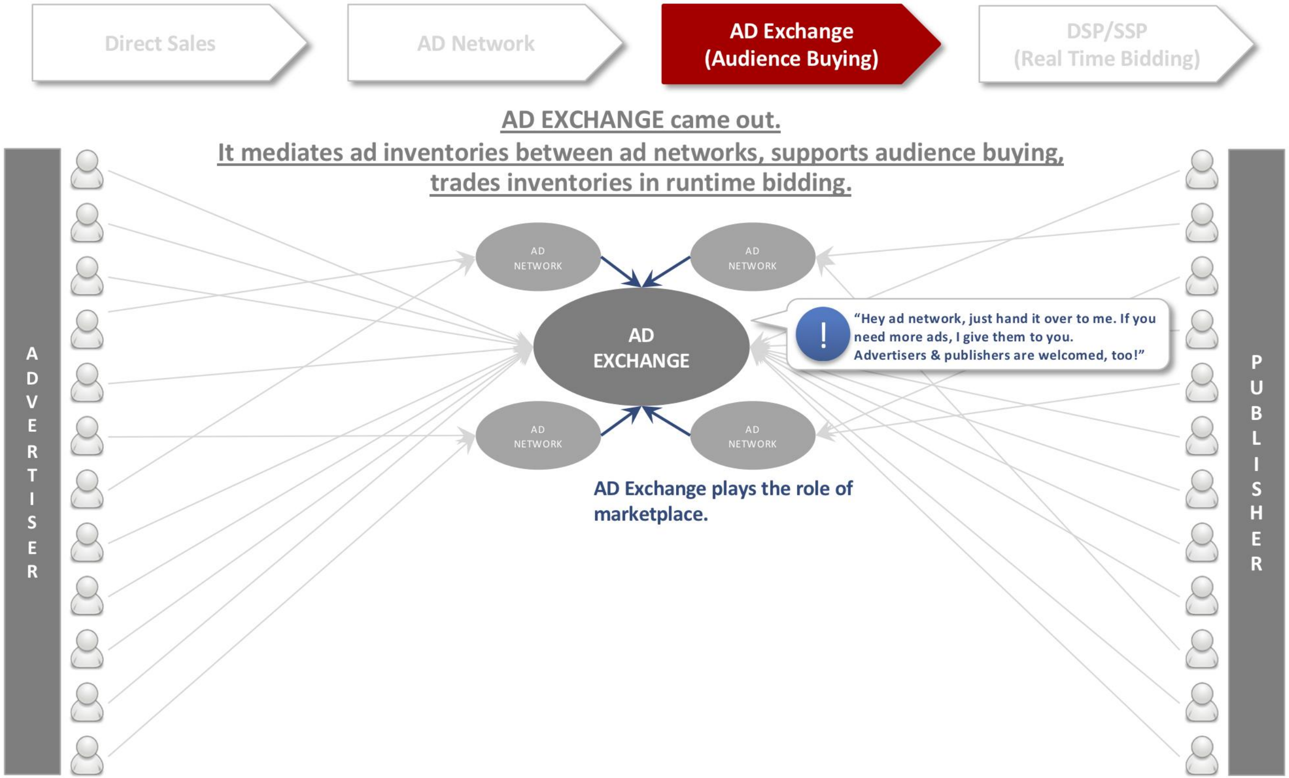
Evolution of digital ad platform: ad network (2/3)



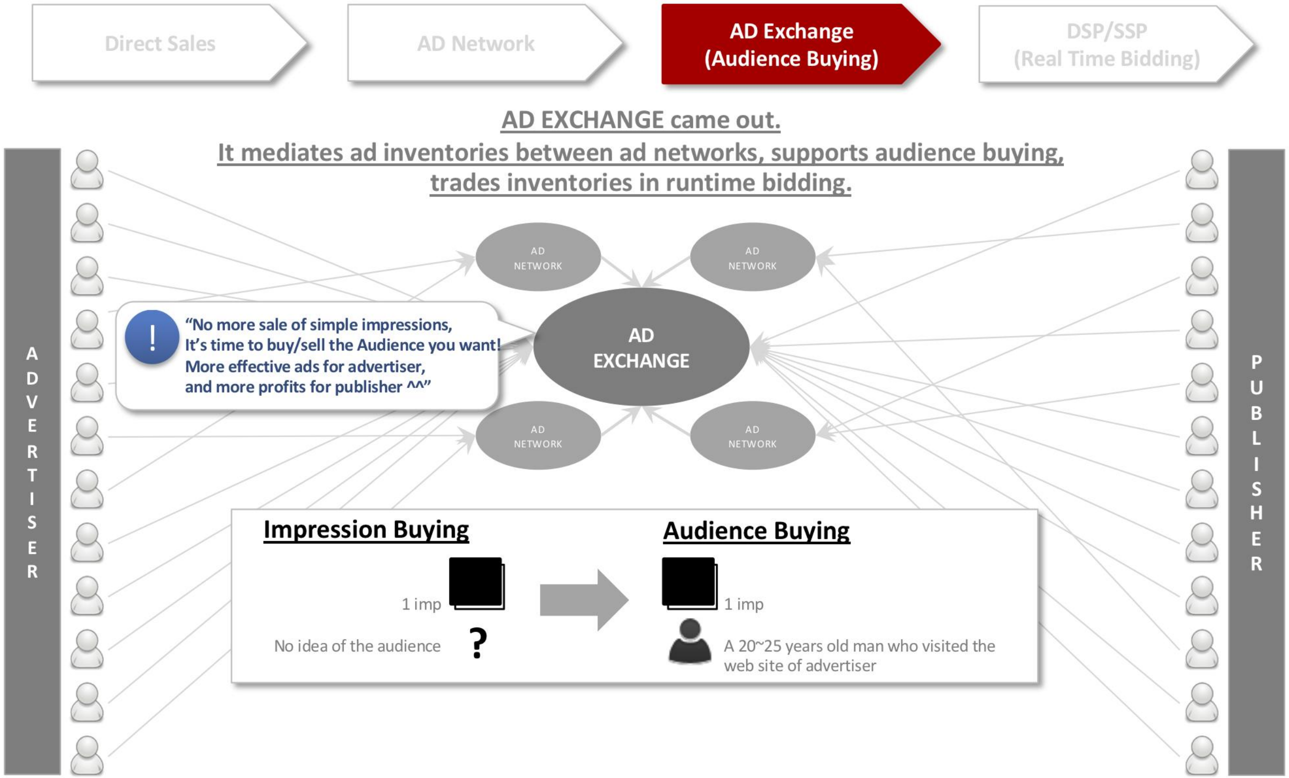
Evolution of digital ad platform: ad network (3/3)



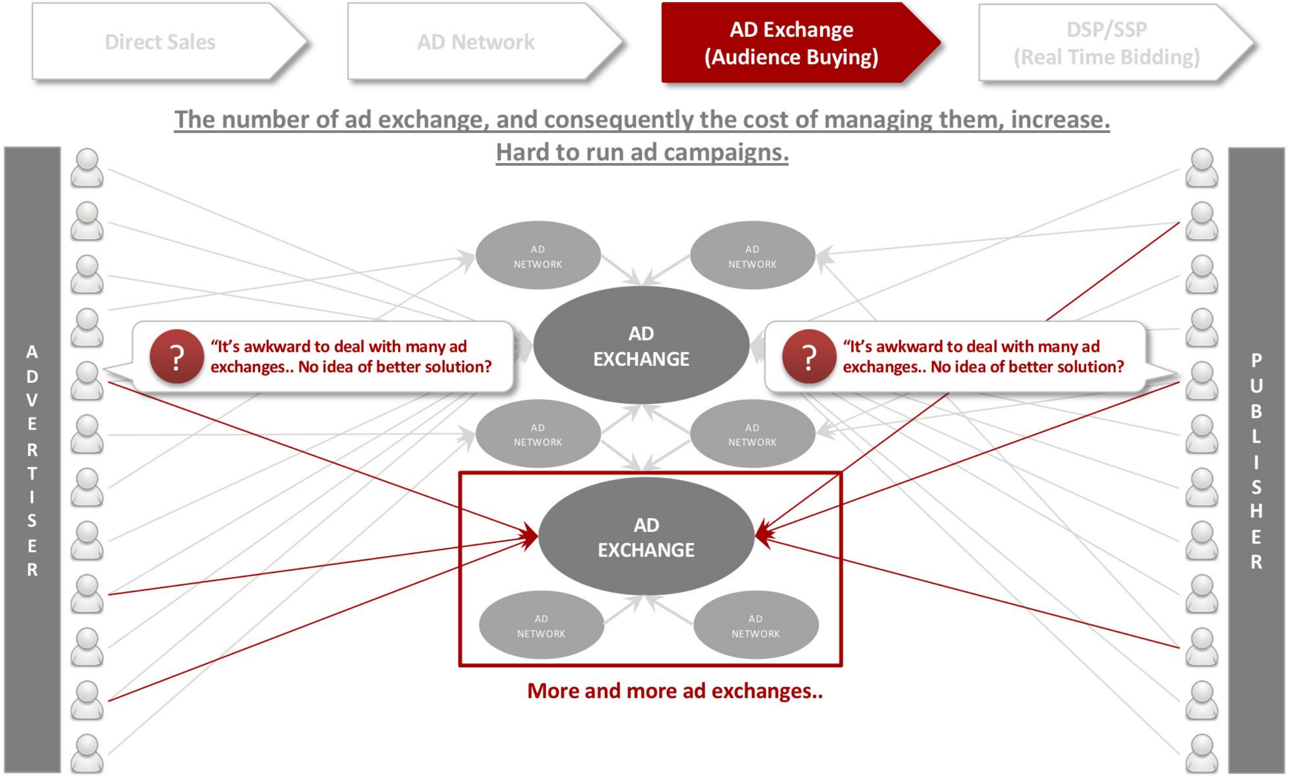
Evolution of digital ad platform: ad exchange (1/3)



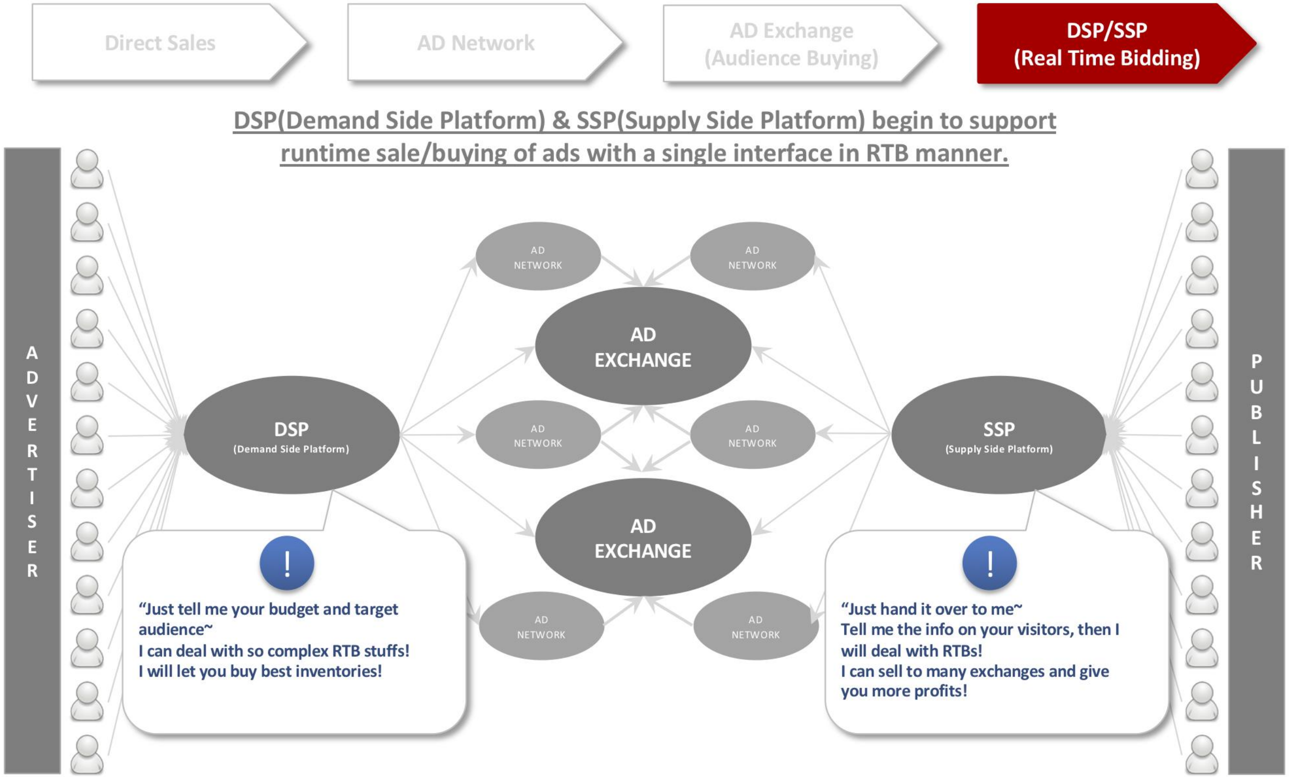
Evolution of digital ad platform: ad exchange (2/3)



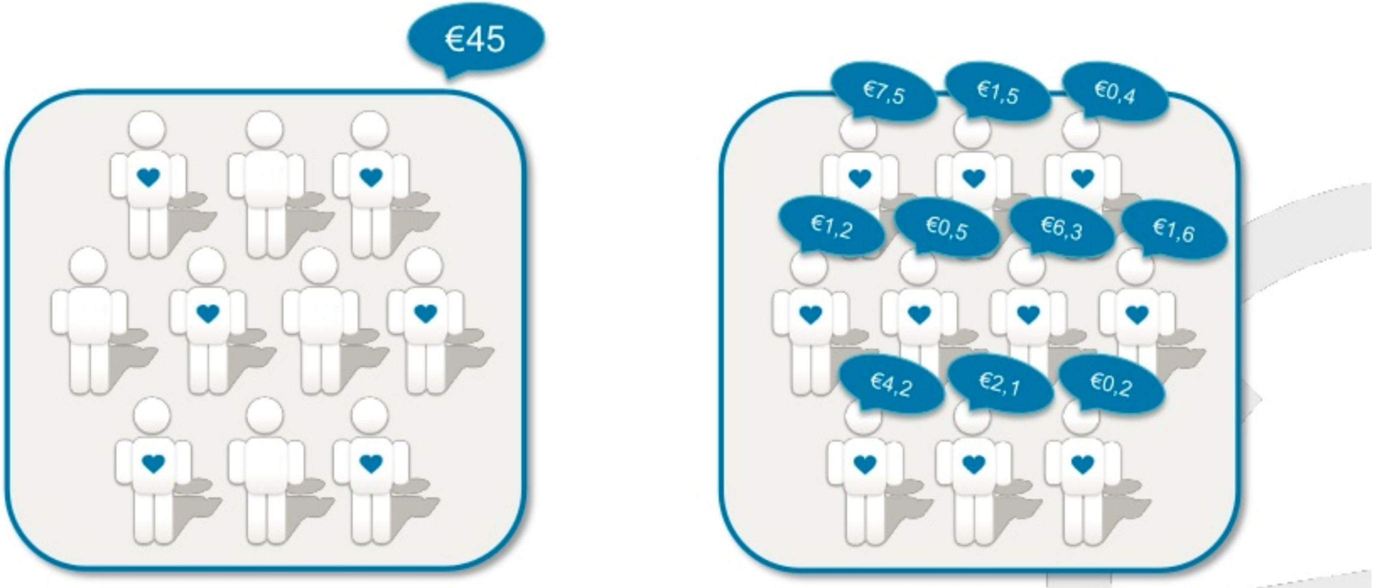
Evolution of digital ad platform: ad exchange (3/3)



Evolution of digital ad platform: DSP and SSP



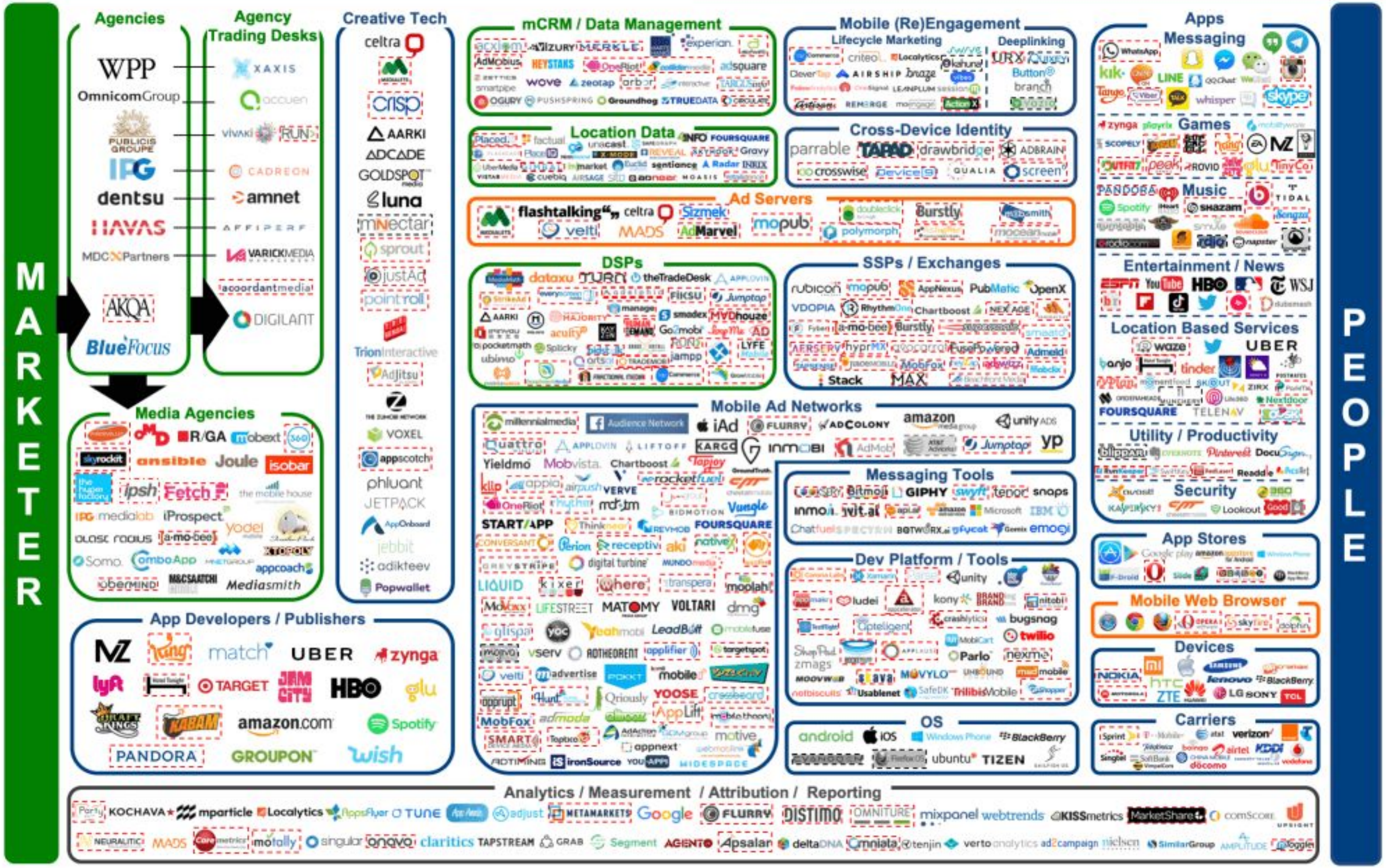
Why Real-Time Bidding: Granularity Matters



Non-RTB:
Impressions purchased in bulk

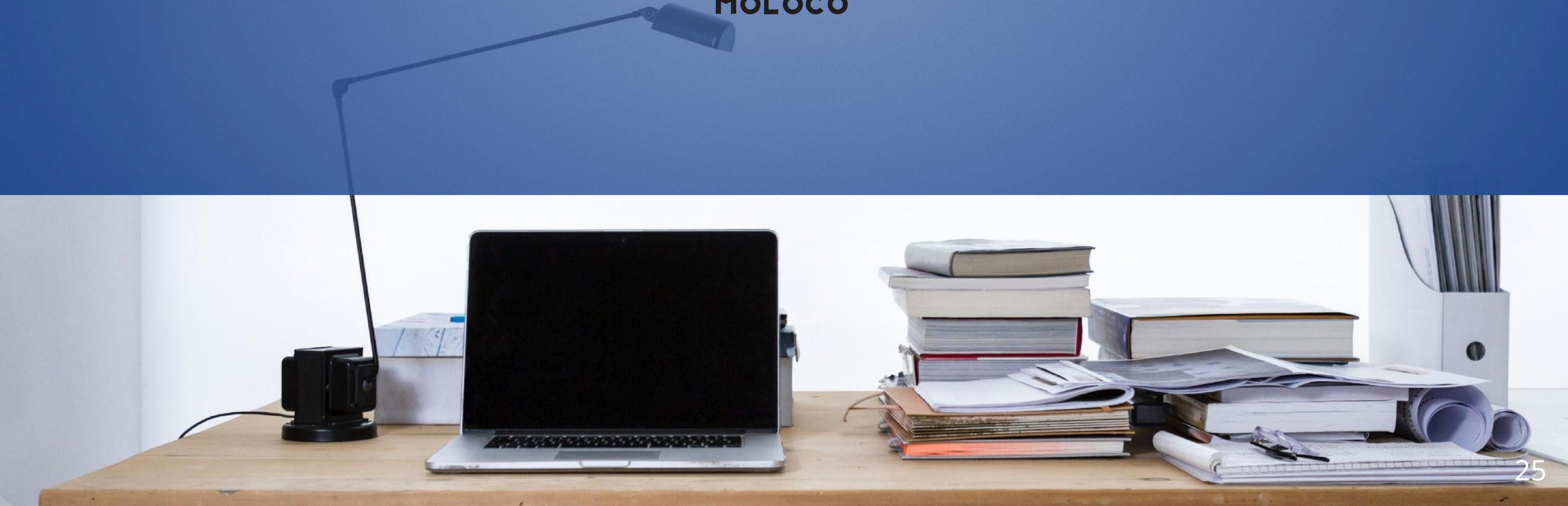
Real-time Bidding:
Each impression won has unique purchase value

MOBILE LUMAscape



Denotes acquired company
Denotes shuttered company

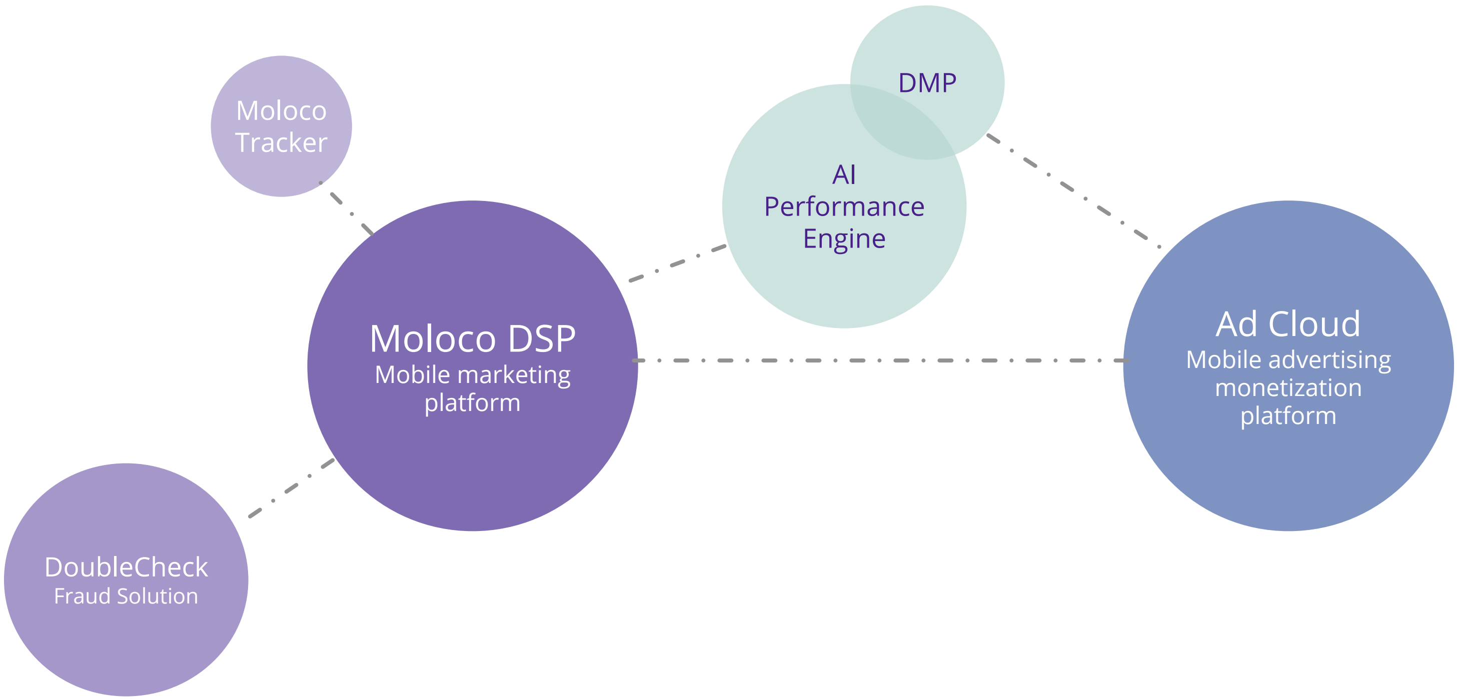
A few things about Moloco



Moloco Offices

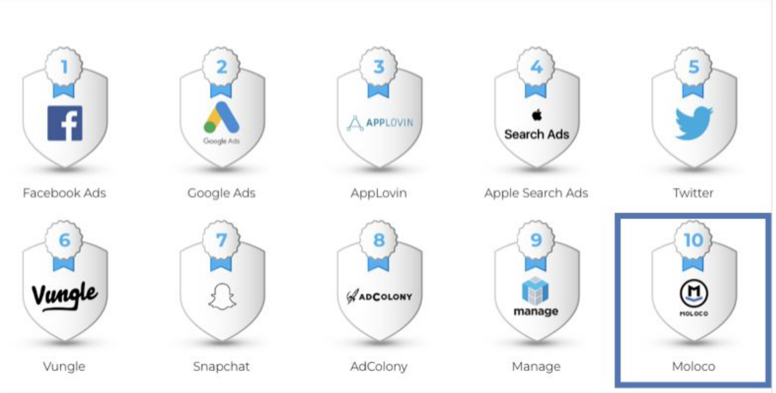
Started in Palo Alto and working from 8 remote offices

Moloco Products



Moloco performance recognition

AppsFlyer Performance Index Power-ranking Top 10 & Kochava Traffic Index APAC & North America #1



Evaluation based on install volume,
user quality and user in-app performance

Ranking based on signal clarity,
fraud rate and user quality

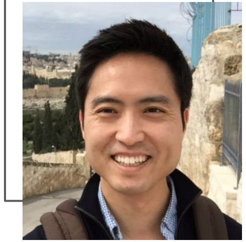
* AppsFlyer Index: https://www.appsflyer.com/2018indexpage/

* Kochava Index: https://lnkd.in/gmVHFQm

Moloco leadership

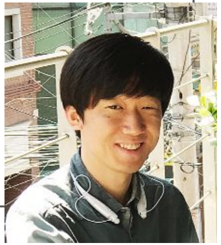
Ikjin Ahn
CEO, Co-founder
Google Android Big data
Tech Lead Manager

Sehyuk Park
Co-founder
Oracle Software
Engineer



Brian Yoo
COO
Google Analyst

Donghwan Jeon CTO
Google Cloud Platform
Senior Software
Engineer



Hyeonseo Ku
APAC Director
Amazon Payment
Software Dev
Engineer

Moloco people are from...

Education background of Moloco people

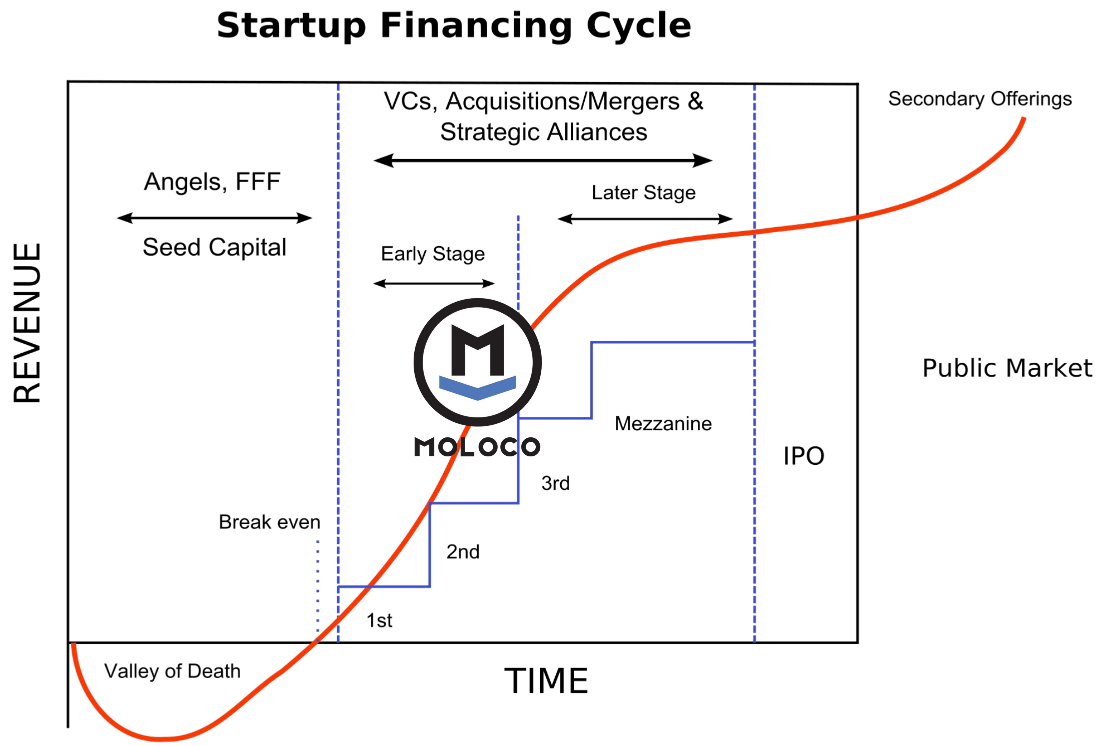
Ph.D.

16

M.S. (incl. MBA)

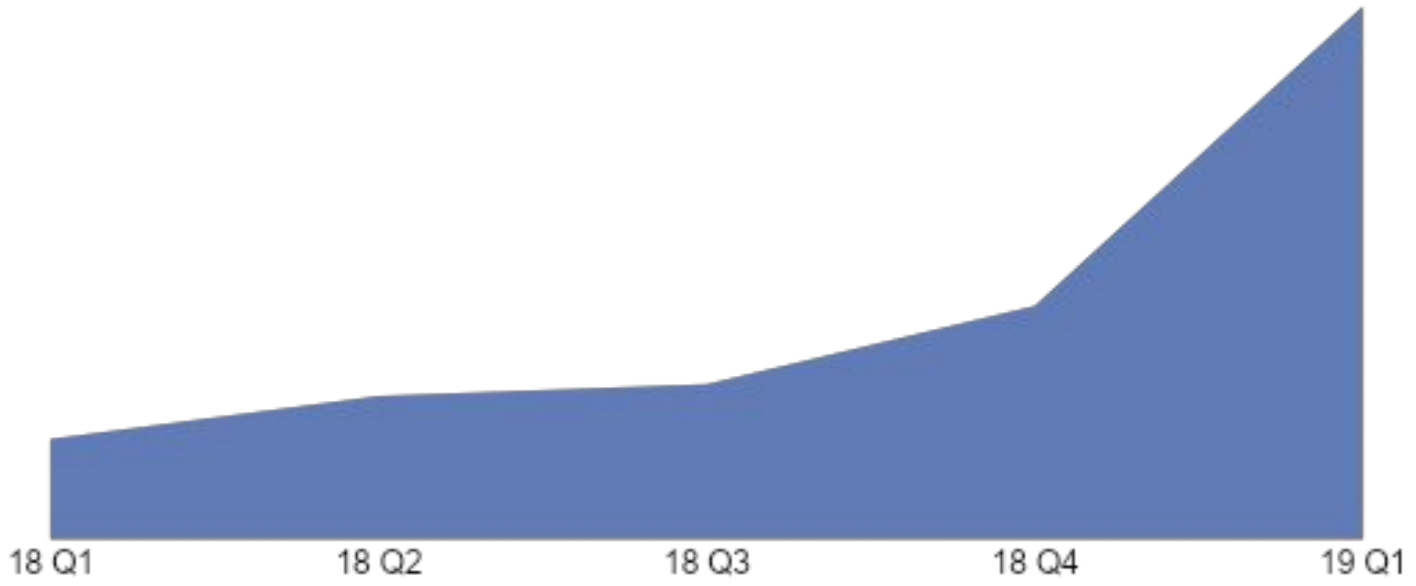
23

Growth of Moloco

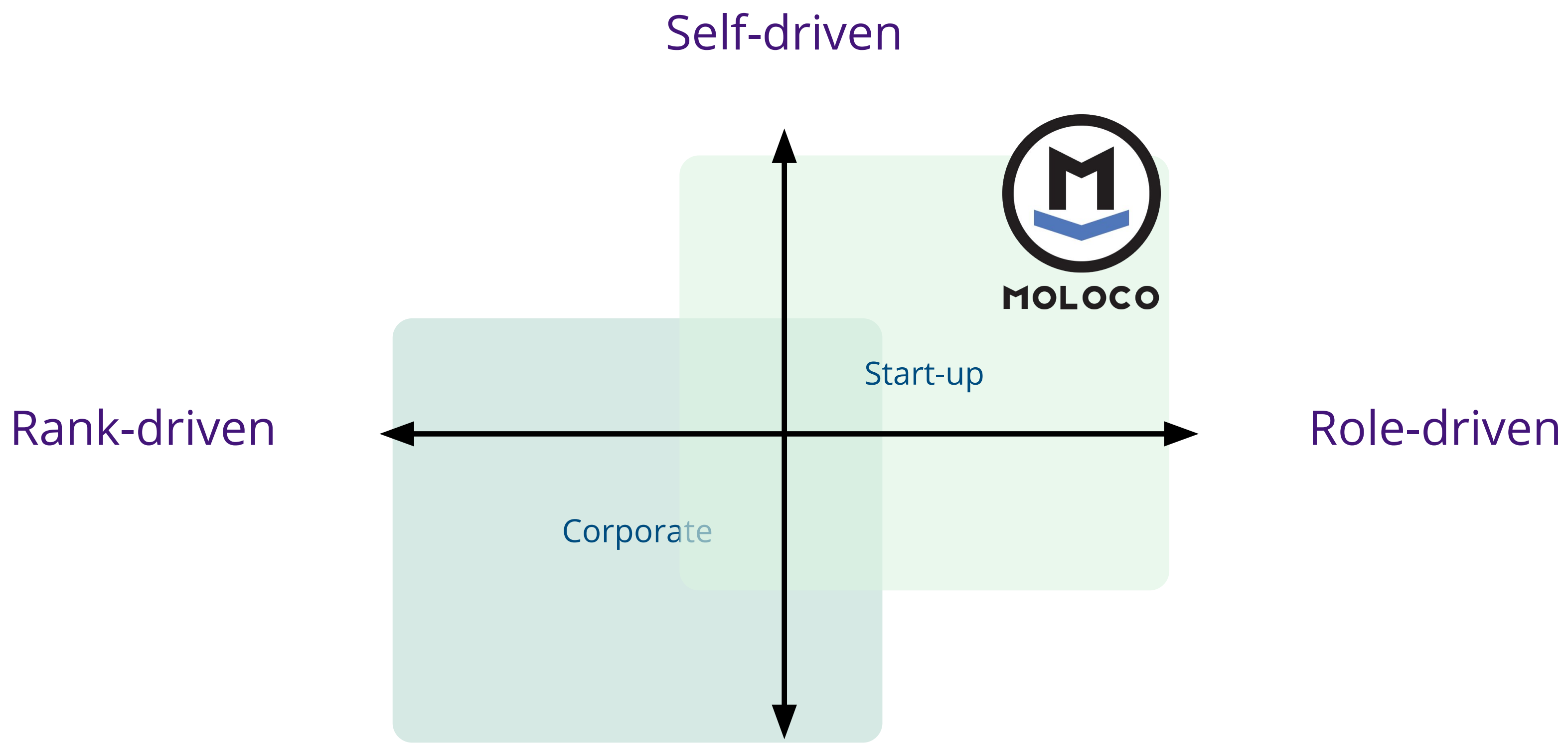


Growth of Moloco

Quarterly revenue of Moloco

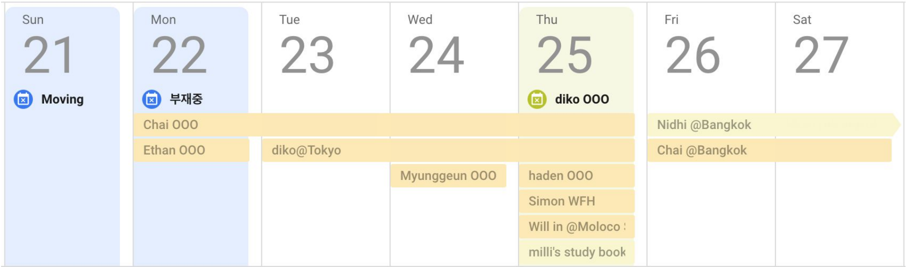


Working style in Moloco

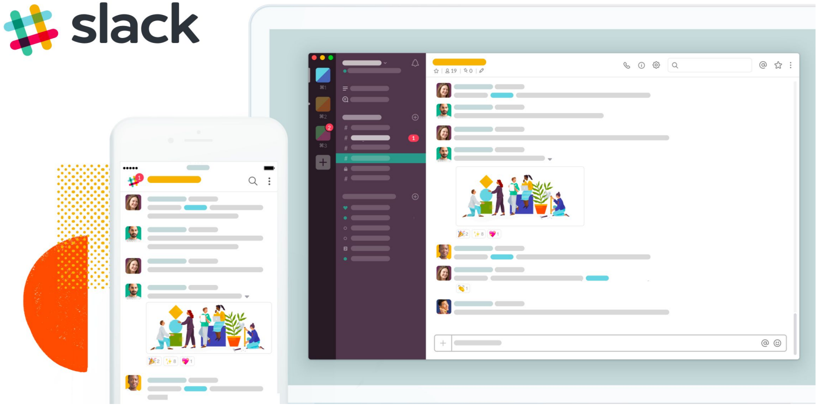


How Moloco people work

Google Calendar



slack



zoom

Moloco people are...

Autonomous

Excellent in problem finding and solving

Obsessive in quality

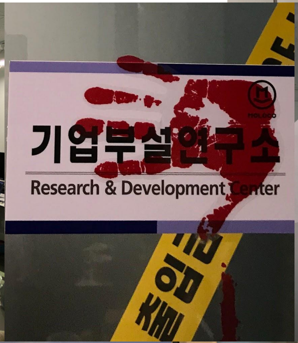
Hard working

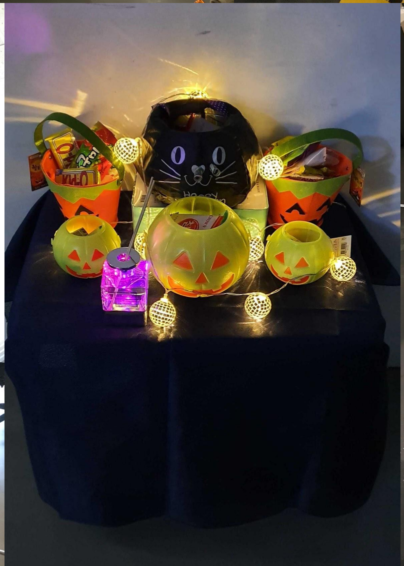
Life at Moloco

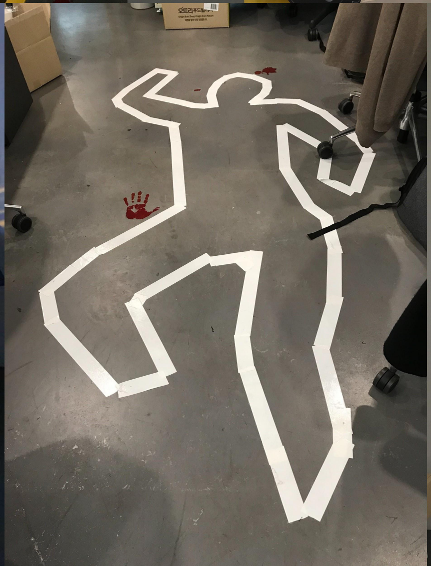
"Moloco 최고의 복지는
가로수길"

2019 Hawaii Global Offsite

Halloween 2019





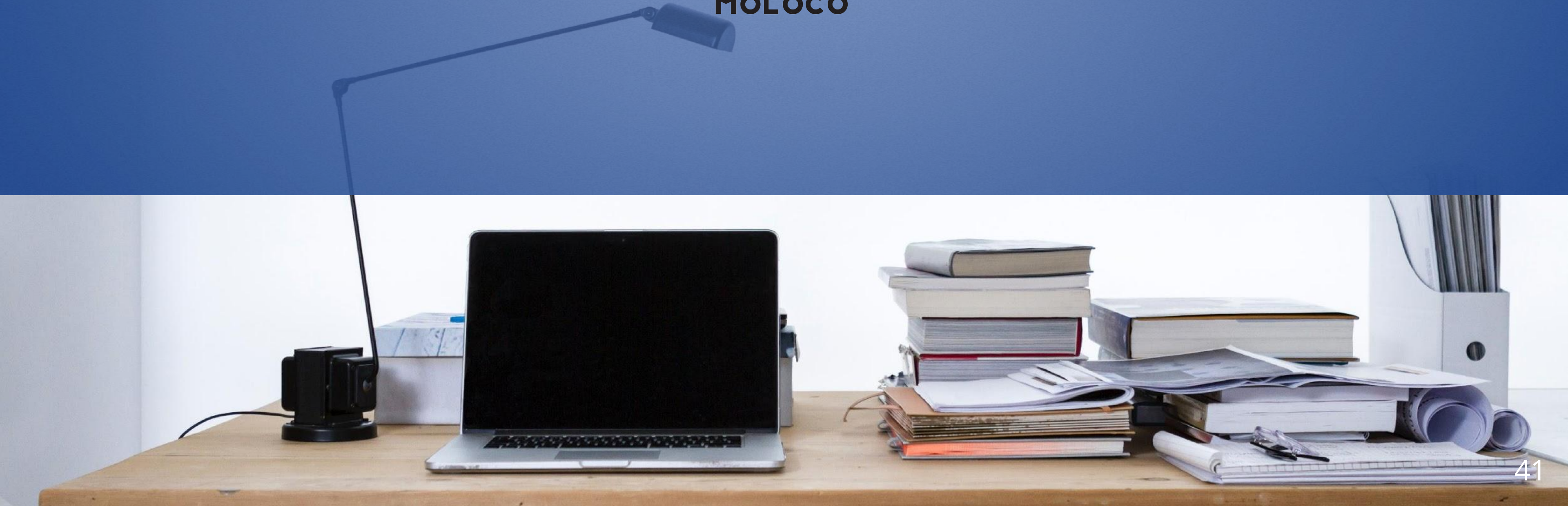


If you want to know more about Moloco...

- https://web.facebook.com/molocoinc/
- https://web.facebook.com/Molocokorea/
- https://medium.com/@moloco
- https://medium.com/@molocokr
- https://www.linkedin.com/company/moloco/
- https://www.linkedin.com/company/moloco/jobs/

How Moloco works as a DSP

MOLOCO

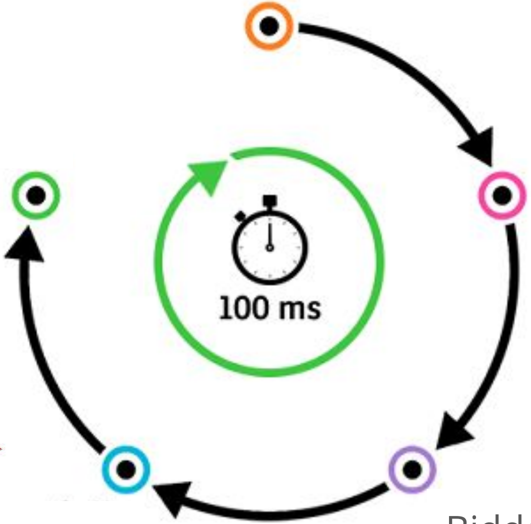


ML models by Moloco

The problem a demand-side platform (DSP) encounters

bidrequest is generated when user opens an app

Winning ad
is served
at user's device



Exchange broadcasts "***bidrequest***" to potential bidders

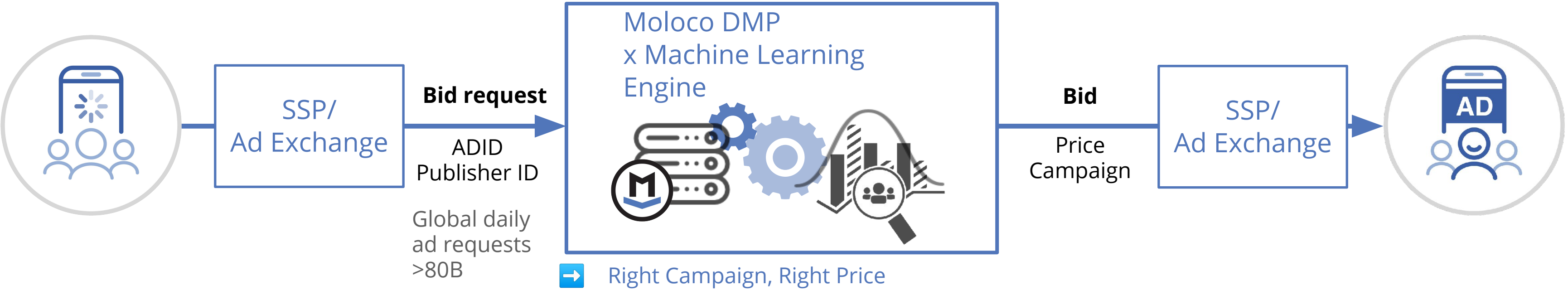
DSP determines the bid price on behalf of advertisers

Auction is over at exchange

Bidders make bids

How Moloco handle this problem with ML models

100% Programmatic & Transparent



➡️ Right Campaign, Right Price

✓ Conversion rate calculated considering Audience x Publisher x Inventory

✓ Campaign KPI and conditions

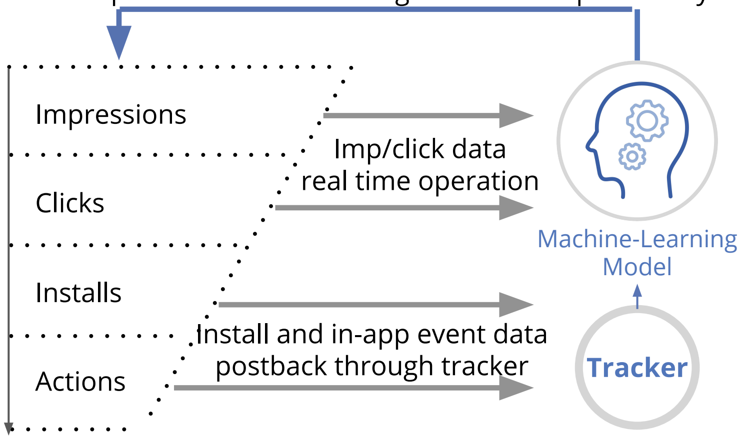
ML engine predicts the user conversion rate after considering factors like user data/publisher/inventory/creatives format

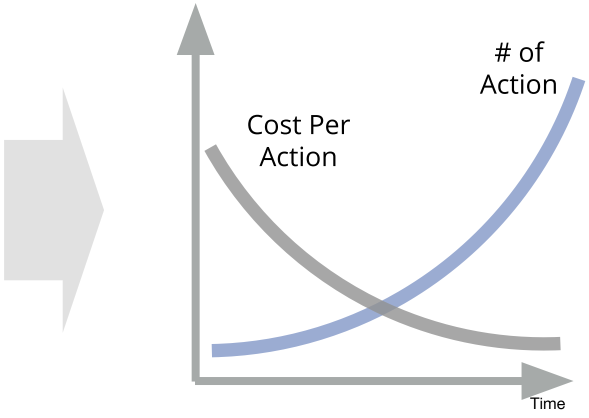
Appropriate price for each ad impression shown

No reselling or re-brokering with Affiliates, only reliable and high quality 100% RTB inventories are secured

Improving performance by machine-learning model

Increasing install prediction accuracy via machine learning
Ad exposure to users with high conversion probability





Optimization with the help of real time postbacks regarding Imp, click, install, in-app actions etc.
Boost the performance by increasing the accuracy of install/action prediction models
(More the data accumulated, easier and faster it is to do the optimization)

Data: sample bid request (1/2)


```
"app": {
    "bundle": "com.fivemobile.thescore",
    "cat": ["IAB12", "IAB17", "news", "sports"],
    "id": "agltb3B1Yi1pbmNyDAsSA0FwcBiq-uwTDA",
    "name": "TheScore - Android App",
    "publisher": {
        "id": "agltb3B1Yi1pbmNyEAsSB0FjY291bnQYpPHvEww",
        "name": "The Score"
    },
    "ver": "6.0.1"
},
```


Publisher App Info


```
"imp": [{
    "banner": {
        "api": [3, 5],
        "battr": [1, 3, 8, 9, 10, 13, 14, 6],
        "btype": [4],
        "h": 50,
        "pos": 1,
        "w": 320
    },
    "bidfloor": 0.970,
    "displaymanager": "mopub",
    "displaymanagerver": "4.14.0",
```


Ad Slot Info
(320x50 banner, at least 0.97 CPM)

Data: sample bid request (2/2)


```
"device": {
    "geo": {
        "city": "Fontana",
        "country": "USA",
        "lat": 34.039214,
        "lon": -117.46634,
        "region": "CA",
        "zip": "92335"
    },
    "ifa": "099c14f1-f3c5-4da2-87b1-XXXXXXXXX",
    "ip": "99.23.172.121",
    "language": "en",
    "make": "LGE",
    "model": "LGMS330",
```


ADID, IDFA

Your Device Info


```
"user": {
   "keywords": "allowHouseAds:false,language:en,GPS:true,league:nfl,
     tab:news,page:article,country:US,
     region:California,city:Fontana,accountType:email,
     articleID:1390672,playerID:/football/players/41055,
     teamID:/football/teams/32,/baseball/teams/26,/football/teams/27,
       /football/teams/676,/basketball/teams/18,
     dmp_LbkziWS8,dmp_LYS180qJ,dmp_LYS1nWJ2,dmp_LV5R4X2g"
}
```


What App Knows about You.
(Please bid higher with confidence!)

Data: sample bid response


```
{
  "id": "1234567890", "bidid": "abc1123", "cur": "USD",
  "seatbid": [
    {
      "seat": "512",
      "bid": [
        {
          "id": "1", "impid": "102", "price": 9.43,
          "nurl": "http://adserver.com/winnotice?impid=102",
          "iurl": "http://adserver.com/pathtosampleimage",
          "adomain": [ "advertiserdomain.com" ],
          "cid": "campaign111",
          "crid": "creative112",
          "attr": [ 1, 2, 3, 4, 5, 6, 7, 12 ]
        }
      ]
    }
  ]
}
```


bidding price (CostPerMillis) :
That is, **$0.00943**

campaign, creative choice

Find the optimal!
In real time, fast enough.

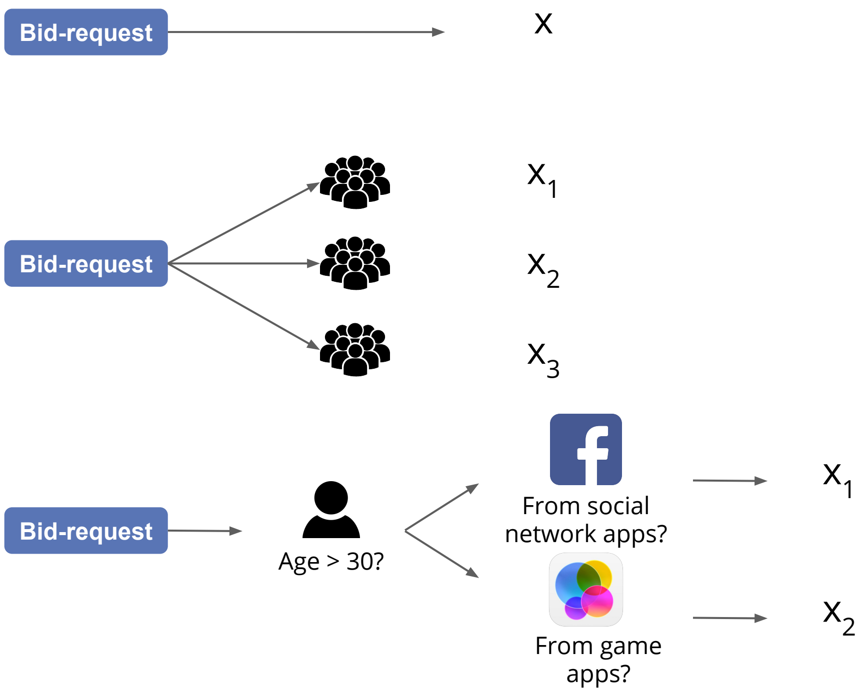
(1) How to make it happen?

(2) How to compute the price
(i.e., what's the bidding strategy)?

What pricing strategy a DSP can have?

Simple pricing strategy which requires low tech (and high manual configurations)

1. Flat pricing
2. Segment-based pricing
3. Rule-based pricing



How do we estimate the value of an impression?

Suppose we run an app-install campaign with payout of $***r*** for each install

Let ***Q*** be the probability that a certain user installs the campaign app after an ad impression, i.e., ***P***(*conversion* | *impression*)

- If the user installs, we receive $***r***
- If the user does not install, we earn nothing
- Expected return = $***r*** * ***Q*** + $0 * (1 - ***Q***)

Value price = reward of target action * target action probability

Thus, we need to build an ML model to predict ***Q***.

Is it Optimal to Bid at Value CPM?

- In 2nd price auction, the **highest bidder wins**, and pays the **second highest bid price**
- Let $\boldsymbol{p_w}$ be the distribution of other bidders' highest bid
- Expected profit of bidding at $\boldsymbol{p}$ is
  - $E[\, (\boldsymbol{r}\, \mathbb{1}\{\text{install}\} - \boldsymbol{p_w})\, \mathbb{1}\{\boldsymbol{p_w} \leq \boldsymbol{p}\}\,]$
  - $= \int_0^{\boldsymbol{p}} (\boldsymbol{r\, Q} - \boldsymbol{x})\, d\mathrm{Prob}(\boldsymbol{p_w} \leq \boldsymbol{x})$
  - $= \int_0^{\boldsymbol{p}} (\boldsymbol{r\, Q} - \boldsymbol{x})\, P(\boldsymbol{p_w} = x)\, dx$
- For any distribution of $\boldsymbol{p_w}$, the expected profit
  - Increases in $\boldsymbol{p}$ for $\boldsymbol{p} \leq \boldsymbol{r\, Q}$
  - Decreases in $\boldsymbol{p}$ for $\boldsymbol{p} > \boldsymbol{r\, Q}$
  - Bidding at $\boldsymbol{r * Q}$ (value CPM) is optimal

How to compute the price?

- (Well-known case) Finding bid-price for a certain ad with following assumption.

1. the client pays us $\$r$ for one click.
2. the second price auction.
3. we know the probability p for the device to make a click from our ad.
4. Market Price (i.e. highest bid-price from others) distribution M is known (with p.d.f f_M).

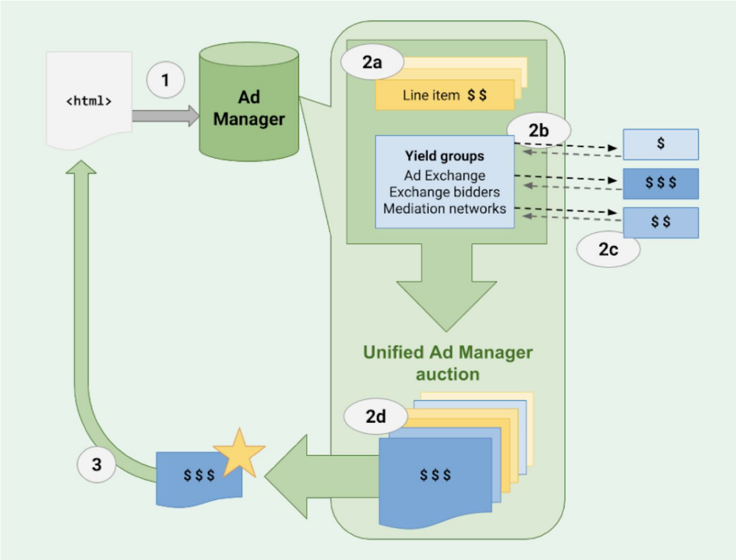
- Then easily the expectation of revenue and profit for bid-price ***b*** is:

$$\mathbb{E}[\text{revenue}](b) = P(M < b)pr$$
$$\mathbb{E}[\text{profit}](b) = P(M < b)pr - P(M < b)\mathbb{E}[M|M < b] = \int_0^b (pr - u) f_M(u) du$$

- If there is no restriction of revenue, profit-maximizing bid-price is $b = pr$
  And this is independent of ***M***.   Generally not this easy!

In Practice...

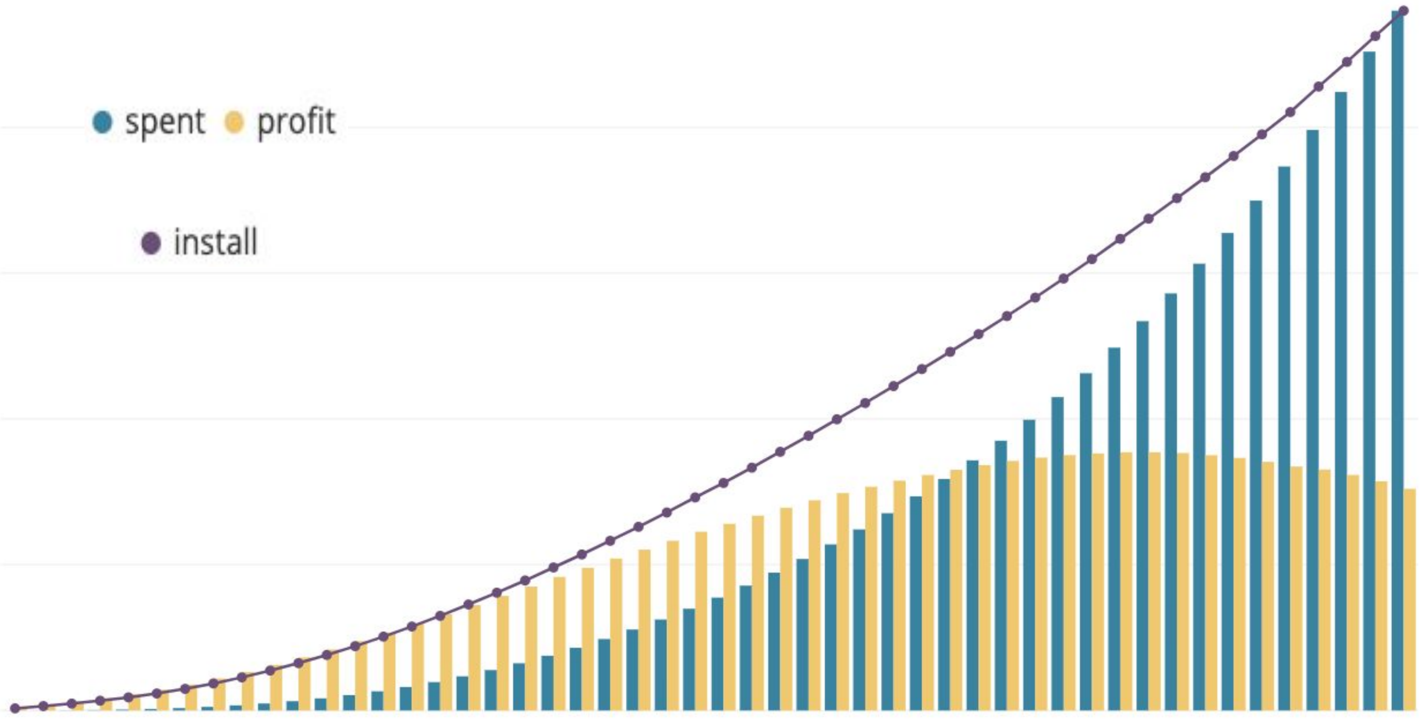
- No exchanges run pure a 2nd-price auction
    - Dynamic floor price (ADX)
    - Multiple bid requests with different floor prices (MOPUB)
    - First-price auction (SAMSUNG)
    - Header bidding & mediation



- Ad-serving cost was NOT considered in the previous analysis

- Control *multipliers* to achieve certain eCPI or volume

How to compute the price? - Realization



Moloco models by campaign status

Lookalike

Seed ADID/
all postback

Install Optimization ML

Moloco-driven Install
events

Action Boosting

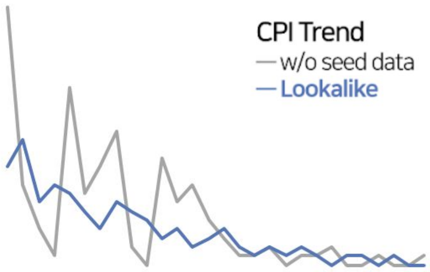
All postback +
Moloco data

Action Optimization DNN

Moloco-driven action
events

As we gather event data (like impressions, click, install, actions, etc) through campaign, ML optimizes the campaign and with enough data, newly matured advance ML model is applied in order to accomplish the campaign goal

Initiating lookalike models



All postback
or Seed ADID

Lookalike Model +
Install Optimization

Performance
Boosting

Best results with the help of healthy seed ADIDs/ all postback settings
Boosting the early stage performance and shorten the period for stabilization via look-a-like model

Logistic Regression (LR) Model

- Logistic regression after getting weighted sum of features and log(feature) values
- Performs relatively well with a small set of data
- Inefficient in exploring new publishers
- Feature engineering is important.
  - Features on publisher, user, campaign
  - Interaction terms, e.g.,
    - publisher X campaign
- *Memorization*, rather than *generalization*
- Can be served in a very fast manner

Campaign	
Cmpgn#1	**0.4**
Cmpgn#2	**0.1**

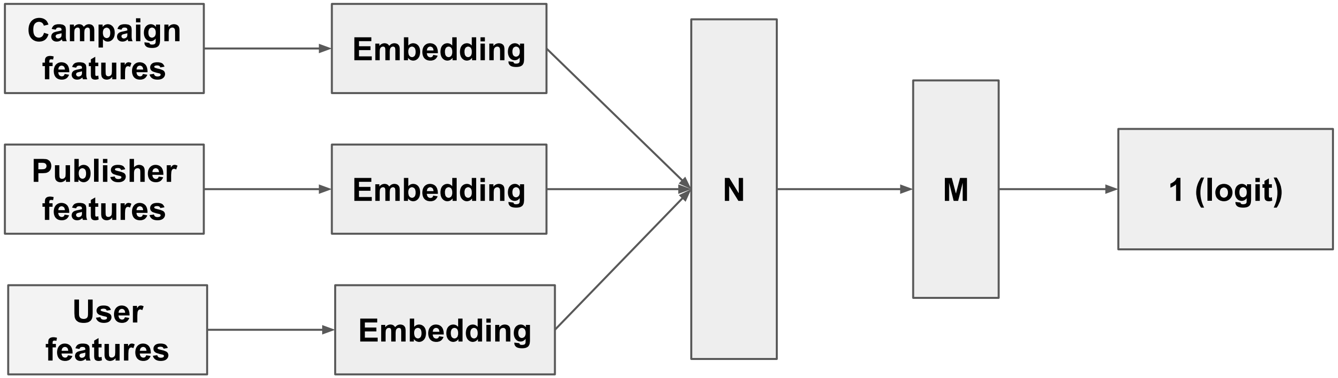
Age	
< 20	**0.3**
>= 20	**-0.1**

Publisher	
APP#1	**0.3**
APP#2	**-0.4**
APP#3	**0.2**

Pub × Campaign	
APP#3 Cmpgn#1	**0.2**
APP#1 Cmpgn#2	**0.3**

Deep neural network (DNN) model

- Reducing repeated computation of matrix is a key.
- Variations of LR/DNN models
  - # of hidden layer, # of hidden unit, layer architecture
  - Selection of positive events: click, install, purchase, etc.



Infrastructure by Moloco

Challenges in building scalable ad system

Small latency with high volume of network traffic

- Real-time bidding (RTB) exchanges require us to respond within 100ms
- We handle >1.6M bid requests per second at peak (>80B per day)

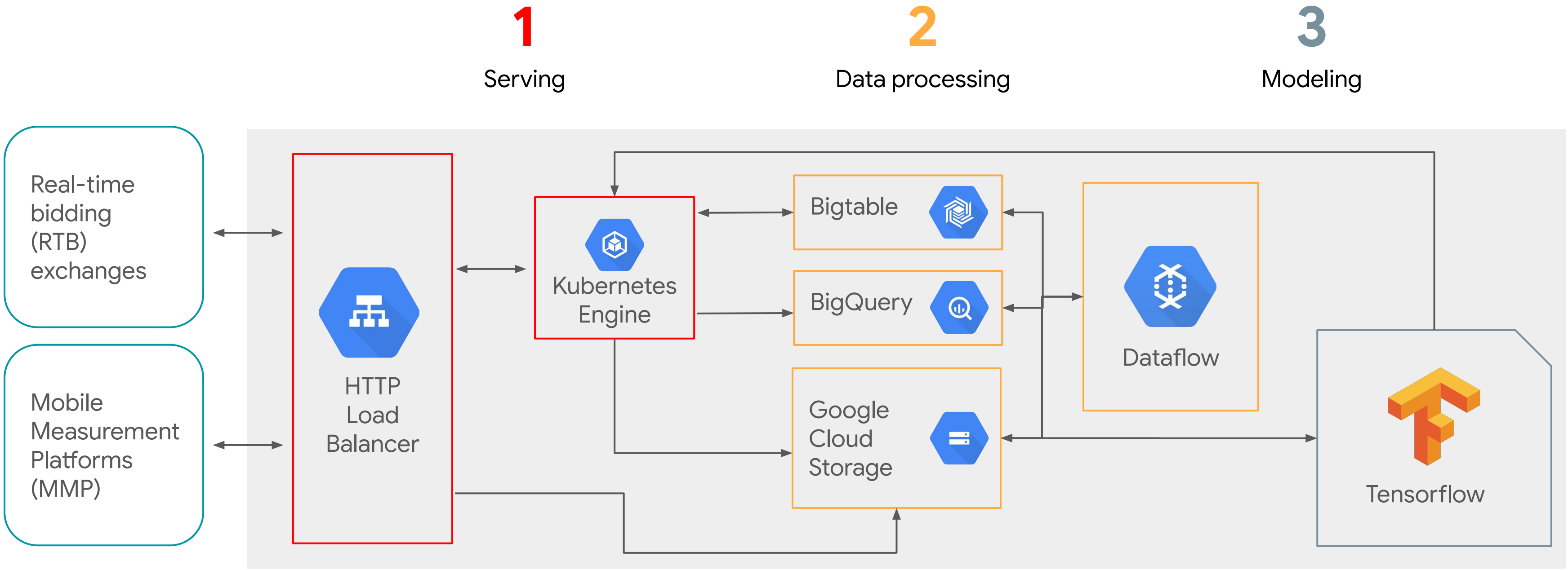
Transforming big data into machine learning models

- >165 TB of new data is generated daily, and we use the accumulated data for training our ML models

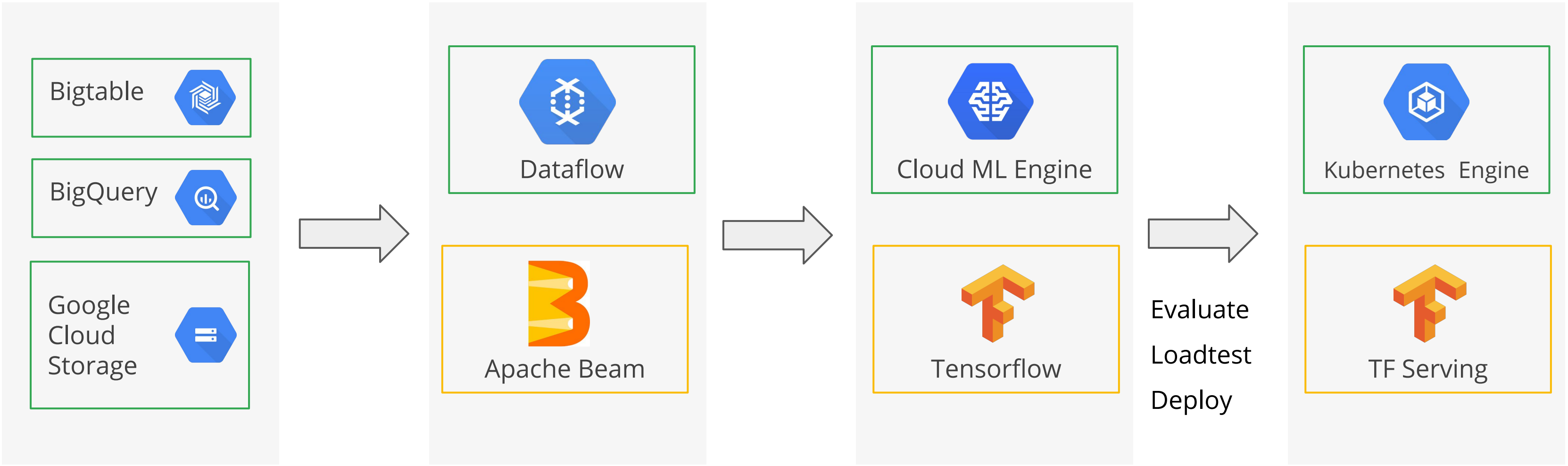
Collaborative tools for complex analysis around big data

- Our data scientists & modelers need collaborative tools to analyze large, complex data

How Moloco is using GCP



Modeling Pipeline



Data

- 5B training examples per week
- 400+ instances

Produce training data as TF examples via python Dataflow

Train Tensorflow model with Cloud ML Engine

- 52 high-CPU instances
- Distributed training

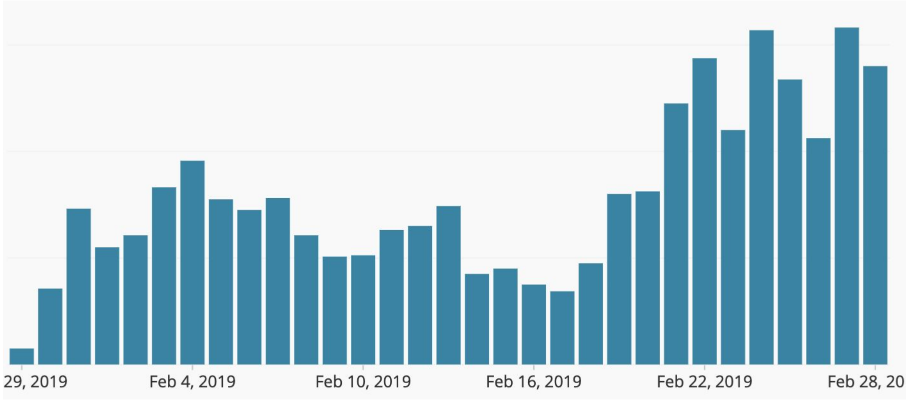
Serve traffic via TF Serving on Compute Engine

- 300+ high-memory instances
- 100+ candidates scored per inference
- 400k+ peak QPS with median latency < 10ms

Example: user acquisition campaign

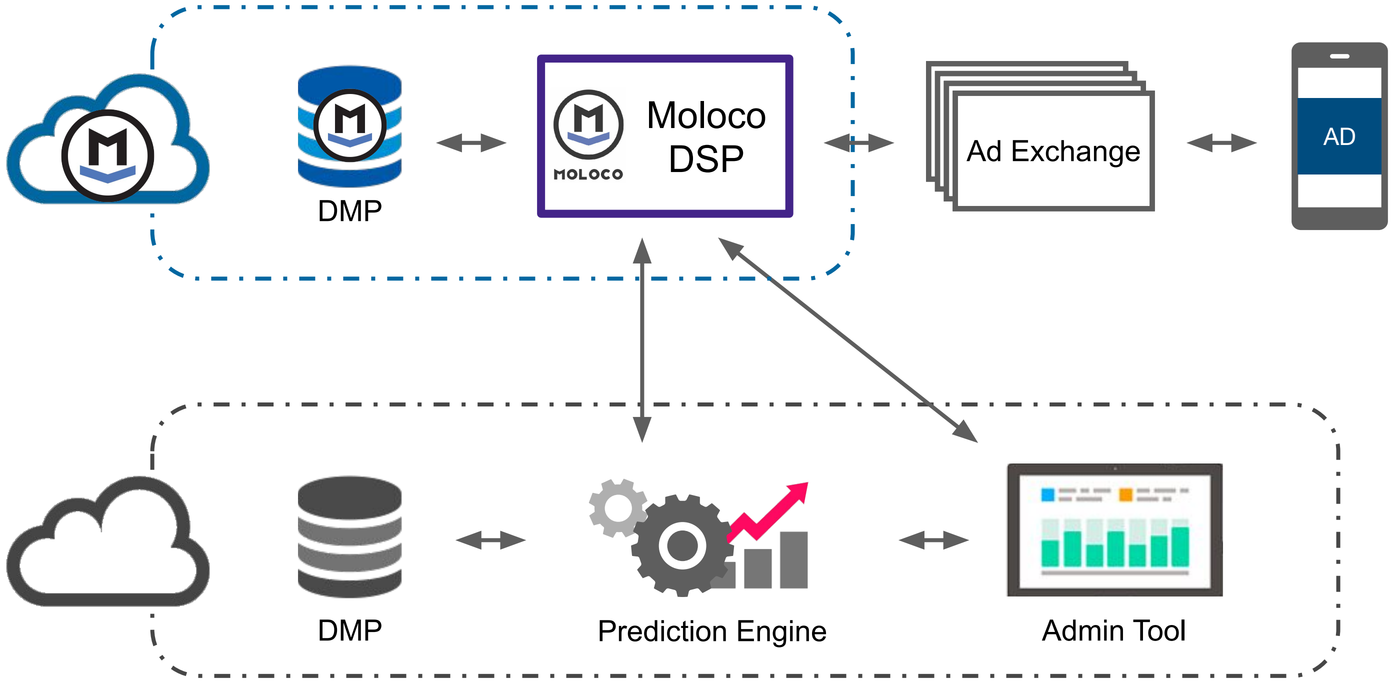
Daily CPI (Cost per Install)

Daily Install



Moloco Ad Cloud

Moloco Ad Cloud enables clients to build custom models using their first-party data



Any Questions...?