

# Big Data in Hulu

Yuming Liang

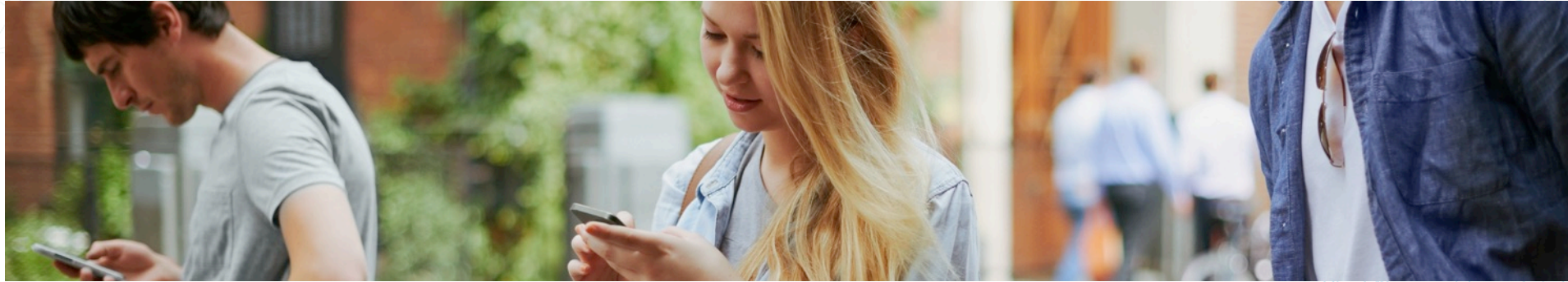
技术架构未来

thegiac.com

# Agenda

- What is Hulu?
- Big Data in Hulu
- How Data Helps Marketing
- How Data Helps Advertising
- Q&A

# Hulu Introduction – Overview



GAMING • CONNECTED TVS • MOBILE • COMPUTER

GIAC | BEIJING  
Dec. 12.16-17

[thegiac.com](http://thegiac.com)



# Hulu Introduction – Overview



# Big Data in Hulu – Framework

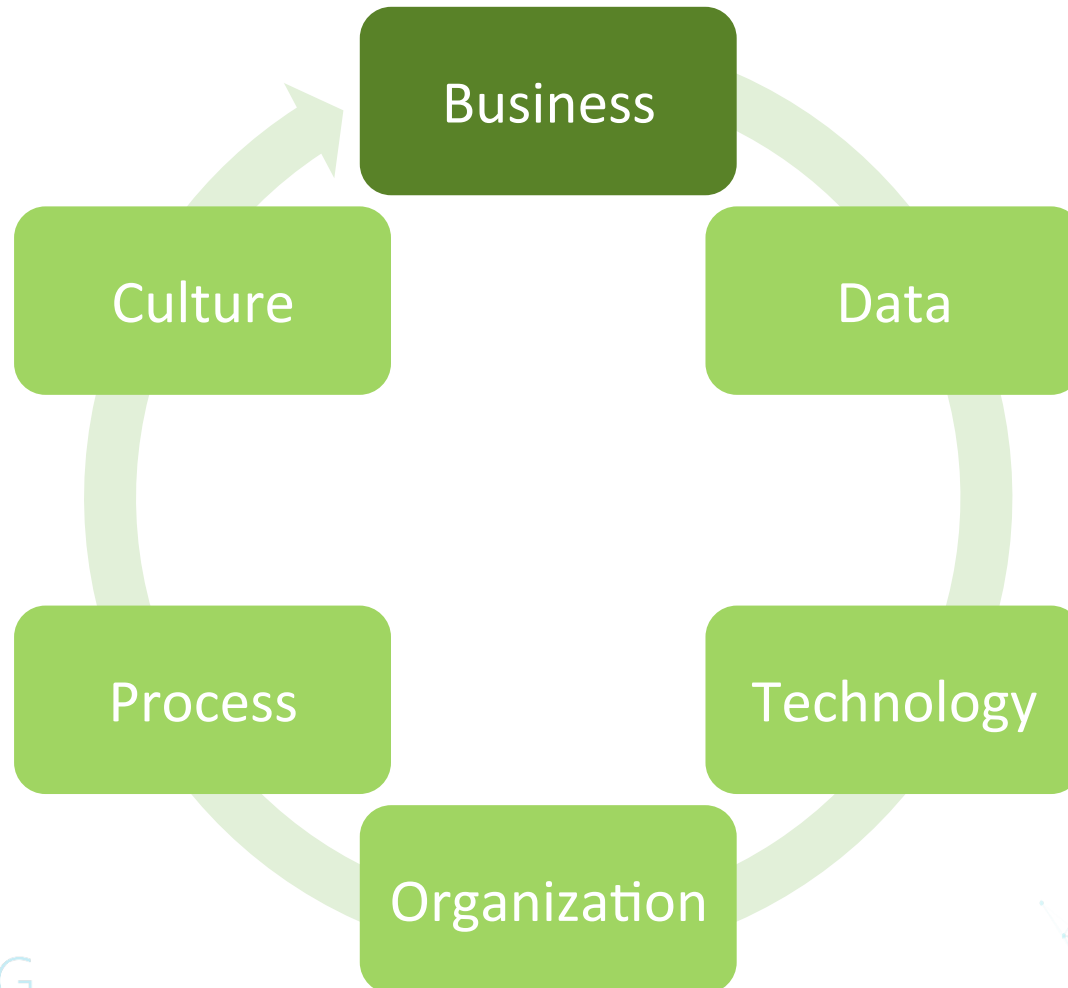


Doing Things Right

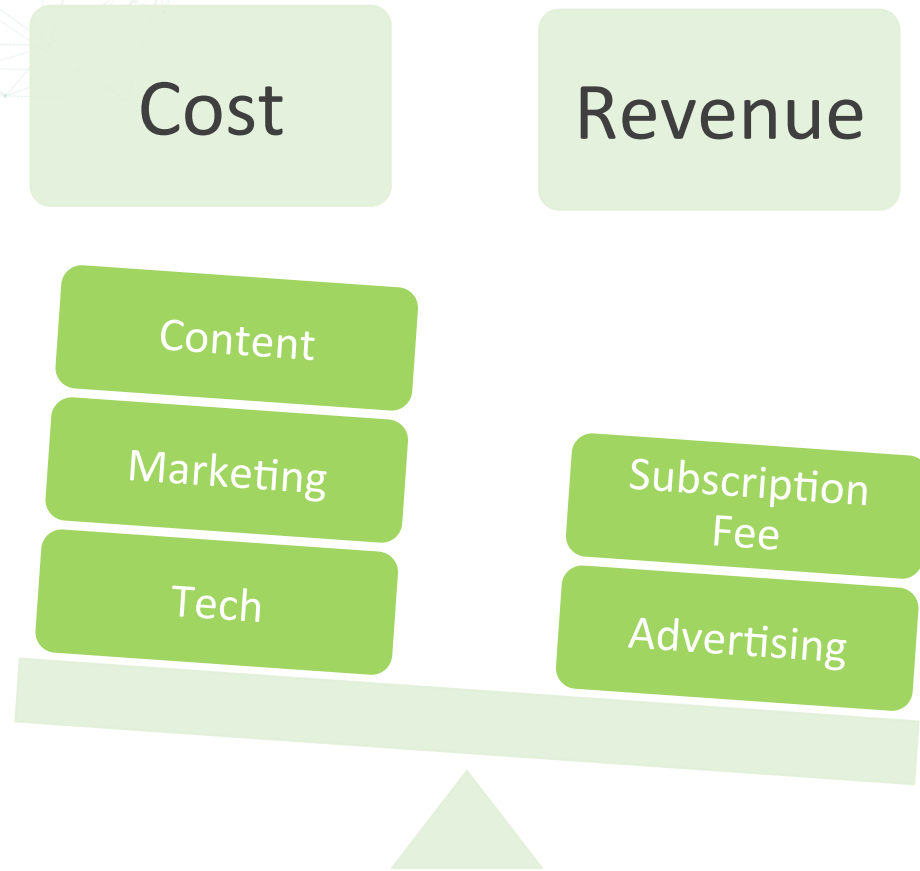
Doing the Right Things



# Big Data in Hulu – Framework

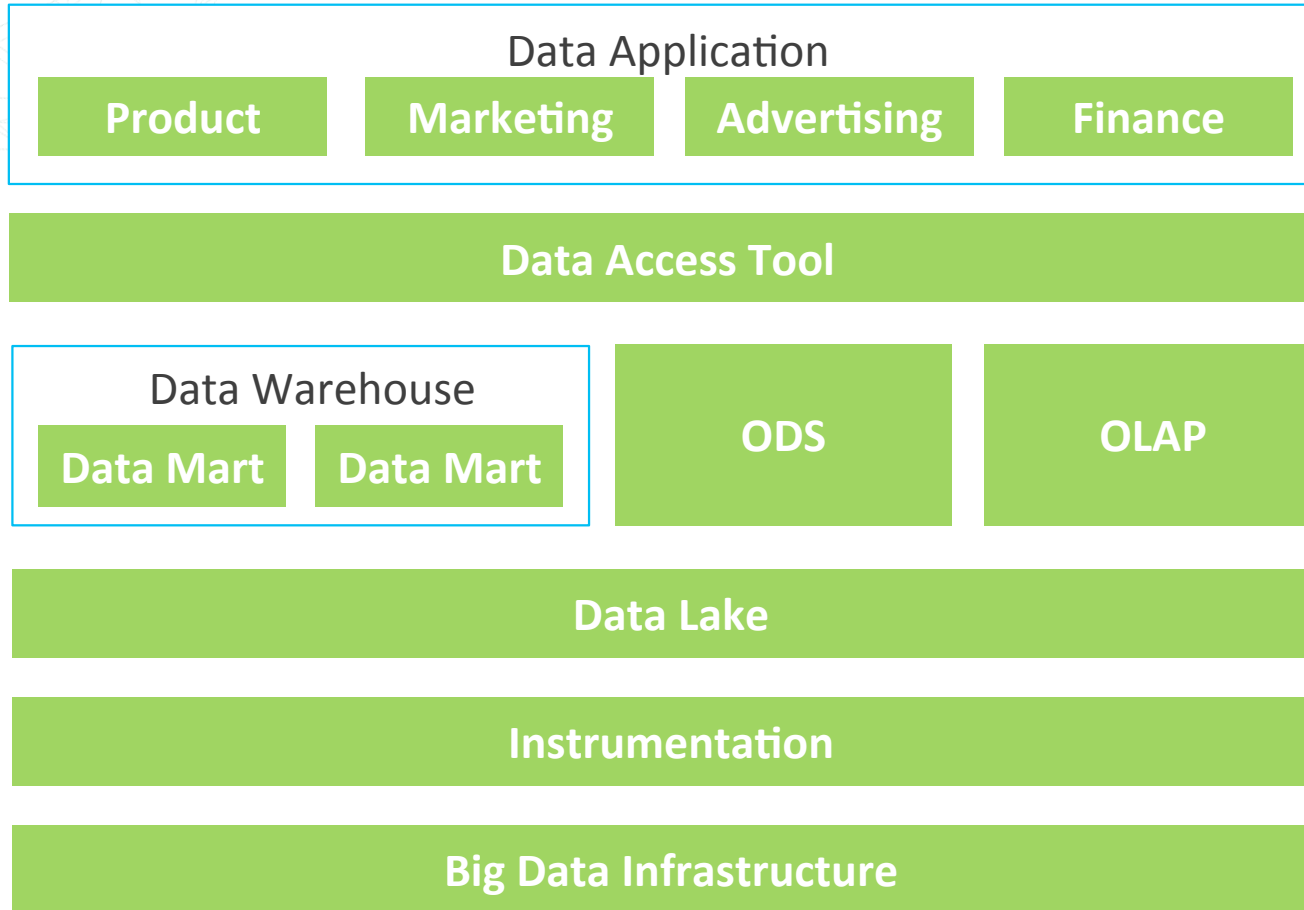


# Big Data In Hulu – Business Model



1. Start from business
2. Figure out how data could help on:
  - support daily operation
  - reduce cost
  - increase revenue

# Big Data In Hulu – Overview



1. Combine general purpose big data technology stack with company specific business scenarios
2. Combine analytics system with operational system



# Marketing in Hulu

# Marketing In Hulu – Goal

Branding

Acquisition

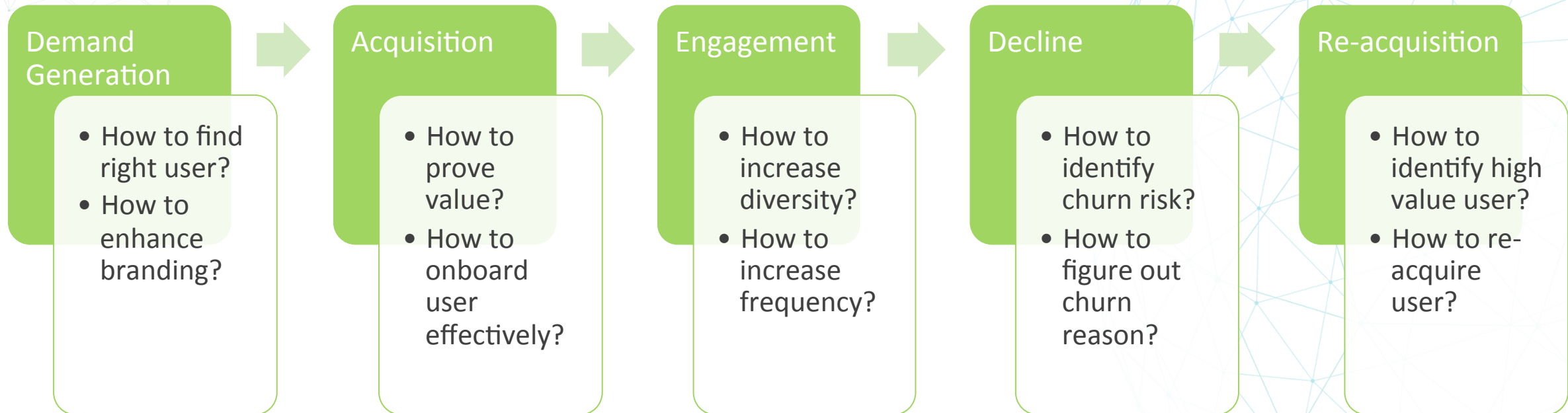
Engagement

Retention



Cost

# Marketing In Hulu – Customer Lifecycle



# Marketing In Hulu – Key Consideration



Context awareness:

1. Varies from different customer lifecycle stage
2. Varies from combination of key consideration

Channel:

1. On Hulu
2. Off Hulu + Online
3. Offline

# Marketing In Hulu – Matrix

	Demand Generation	Acquisition	Engagement	Decline	Re-acquisition
Targeting					
Product					
Pricing					
Channel	What is the ROI for each of our paid media channels? What cost per sign-up should we target to profitably acquire new subscribers? Use CLV and Attribution in combination to determine the CPS target for each channel.				
Messaging					

Data:  
1. Cost / Value of each channel  
2. CLV of each user  
System:  
1. MTA, Dimensional analysis  
2. CRM database



# Marketing In Hulu – Requirement

## Data Lake / Data Mart

- First party / Third party
- KPI / Reporting

## CRM system

- Fact
- User persona

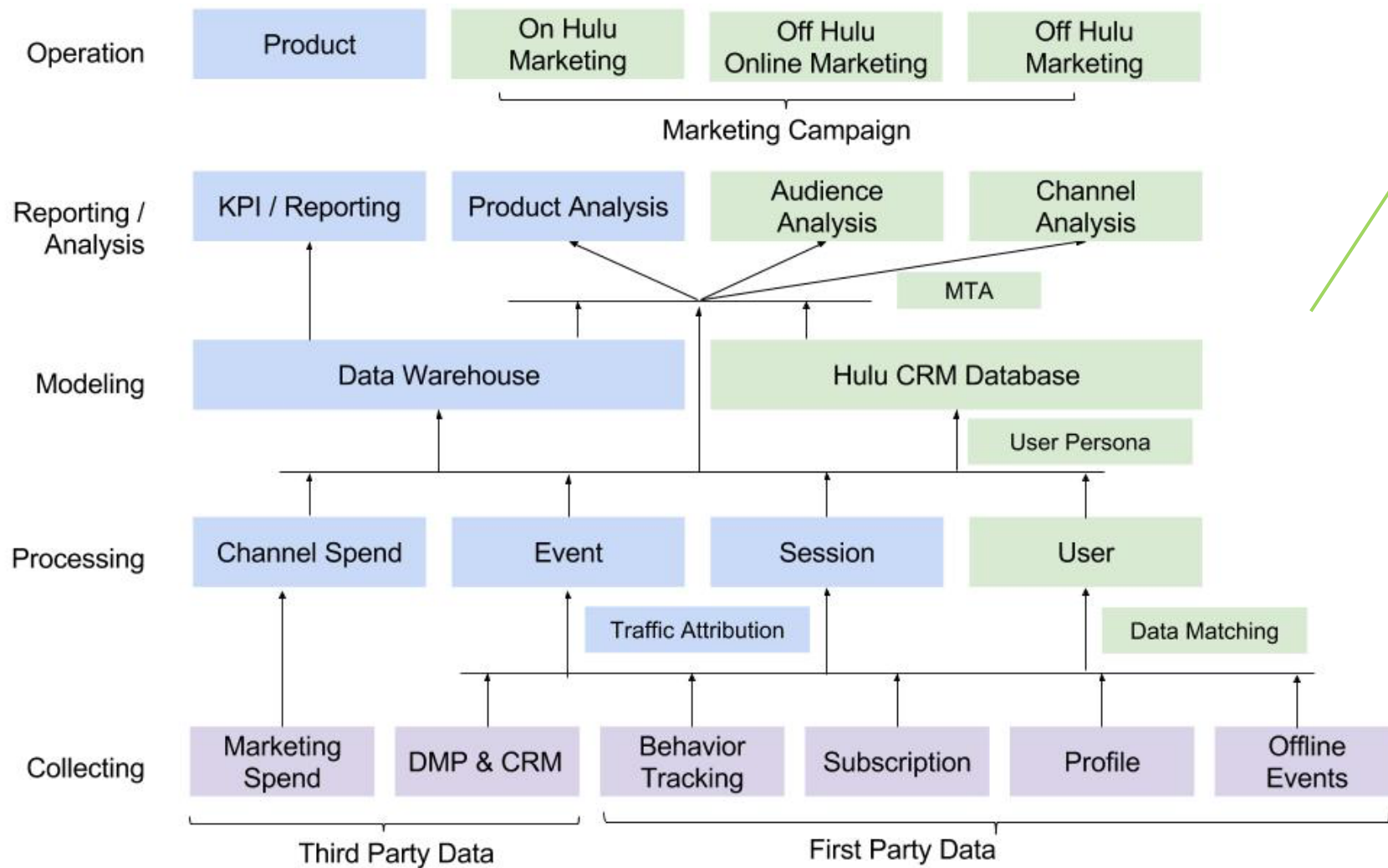
## Marketing campaign management

- In house
- External

## Advanced analytics system

- MMM / MTA
- OLAP / Cohort / Funnel
- Data Mining

# Logical Architecture



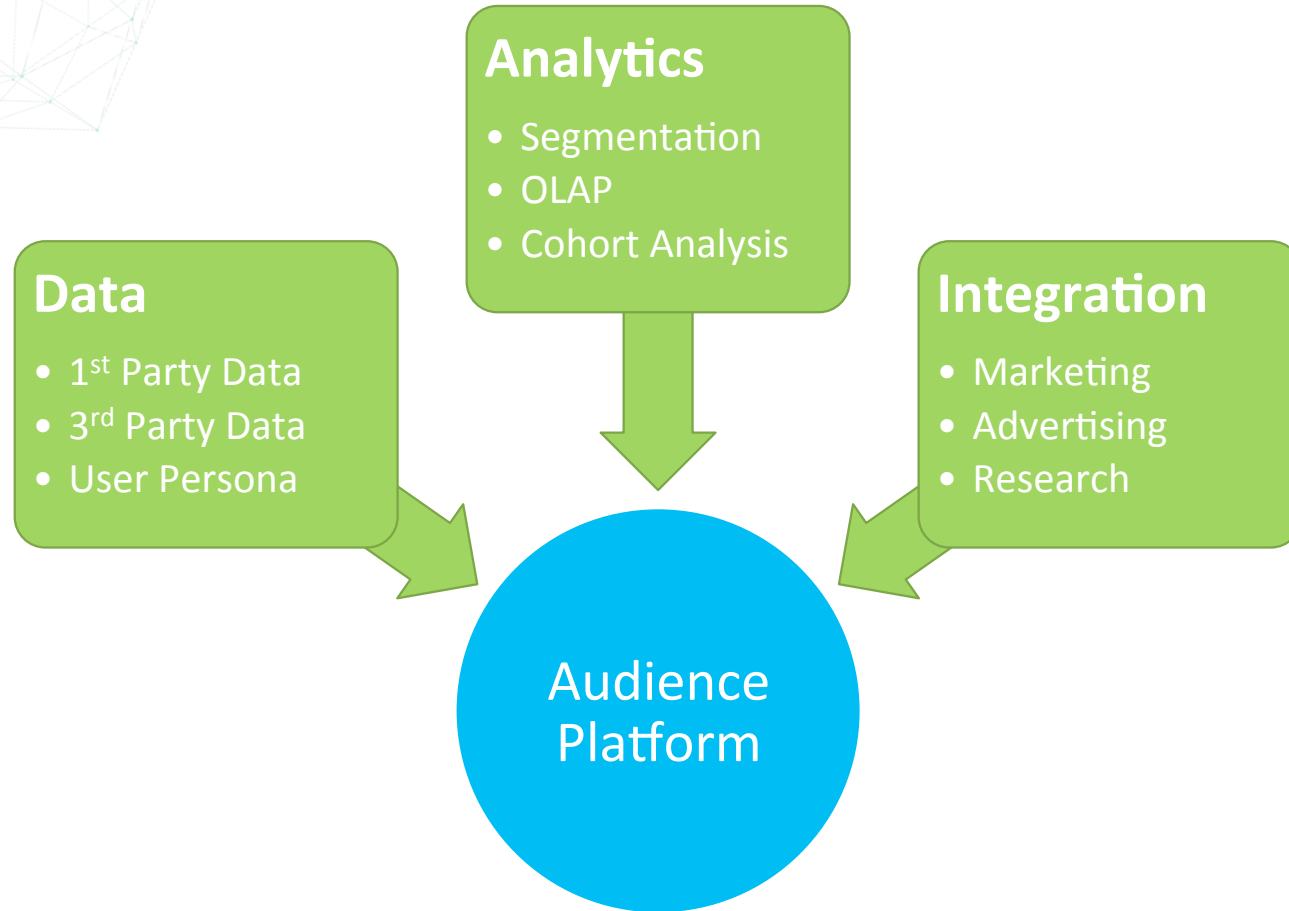
## Data:

1. 1<sup>st</sup> party vs 3<sup>rd</sup> party
2. implicit and explicit

## System:

1. analytics oriented vs operational oriented
2. Fine-grained vs coarse-grained

# Audience Platform - Overview



1. Aggregate data at user level
2. Optimized for operational purpose
3. Optimized for advanced analysis

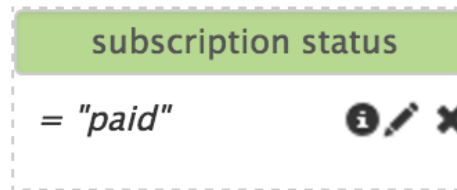
# Audience Platform - Overview

## DATA SOURCE

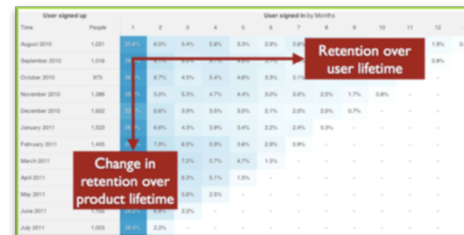
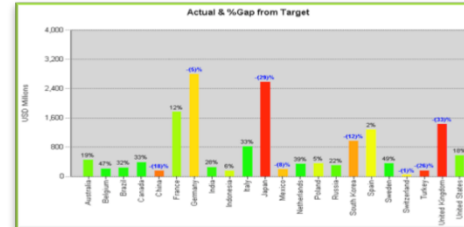
- First party data
  - Profile
  - Behaviors
    - Onboarding
    - Engagement
    - Payment
    - Churn
  - User persona
    - Advertisement
    - Product
    - Marketing
    - Finance
- Third party data
  - Facebook
  - BlueKai
  - Acxiom
  - Experian
  - \*DataLogix

## CORE FUNCTIONS

- User Cohort



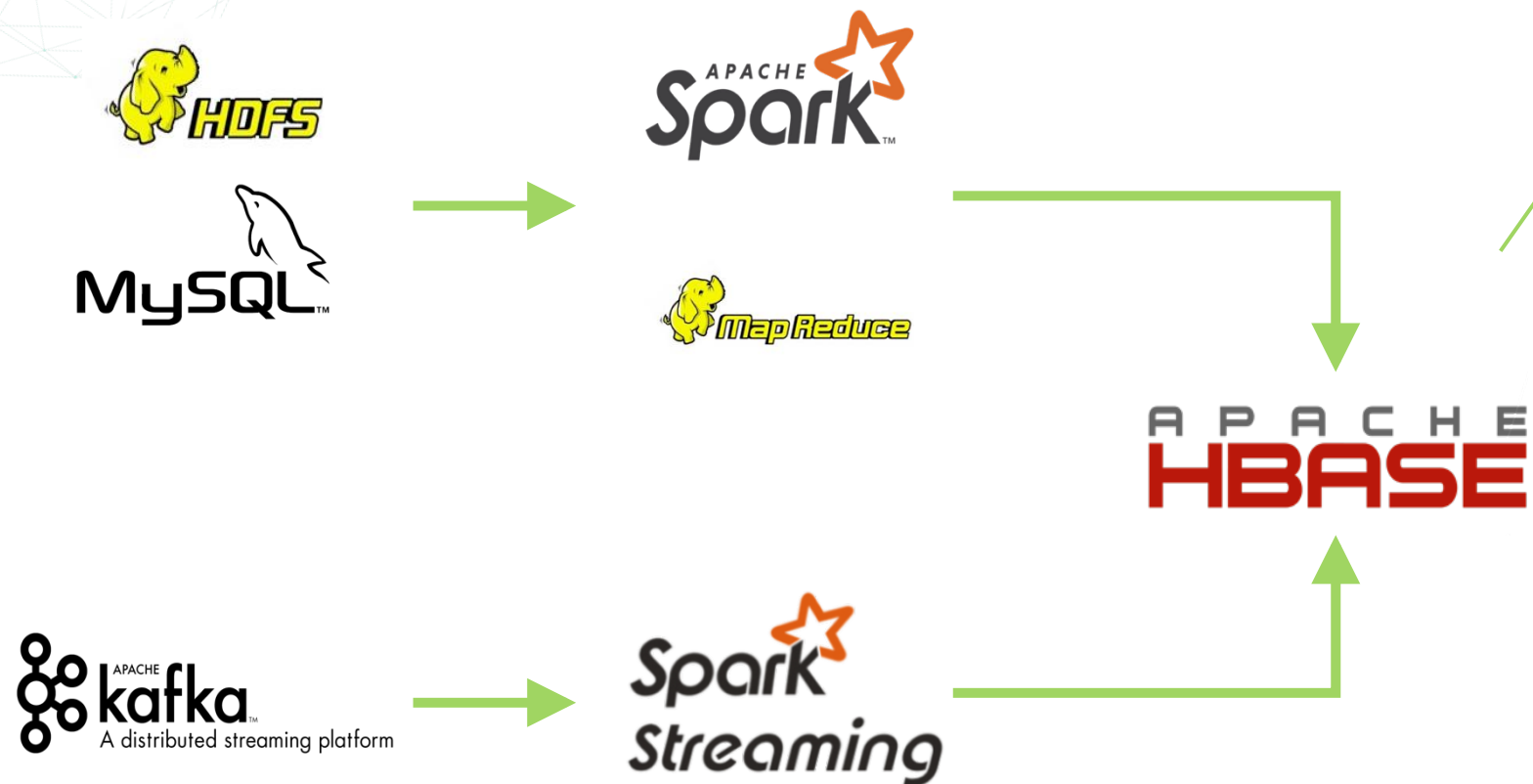
- Deep analysis



## INTEGRATIONS

- Marketing
  - Online + On Hulu
    - Push notification
    - Email
    - In house ads
    - Masthead
  - Online + Off Hulu
    - Social Targeting
    - DMP
  - Offline + Off Hulu
    - Analysis
    - Panel matching
- Advertisement
  - Audience Targeting
    - First party
    - Third party
  - Exchanging data with third party

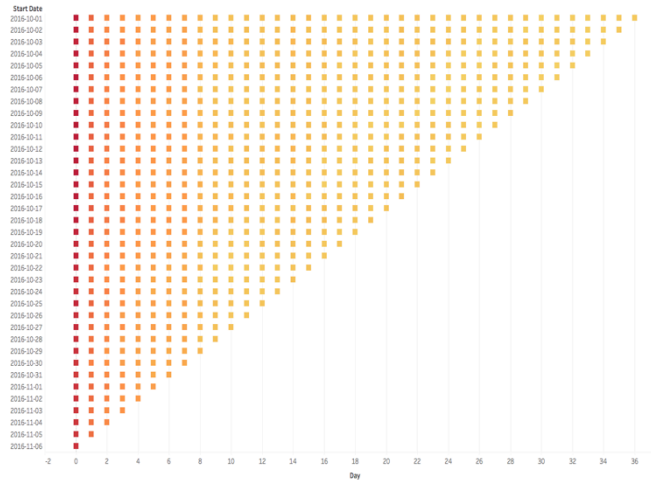
# Audience Platform – Data Pipeline



1. Logical architecture
2. 3<sup>rd</sup> party is ingested in similar way
3. Several down streaming integration



# Audience Platform – Analytics Engine (1)



1. How to make it fast enough?
2. Presto is too slow for this cases due to joining too many tables

# Audience Platform – Analytics Engine (2)

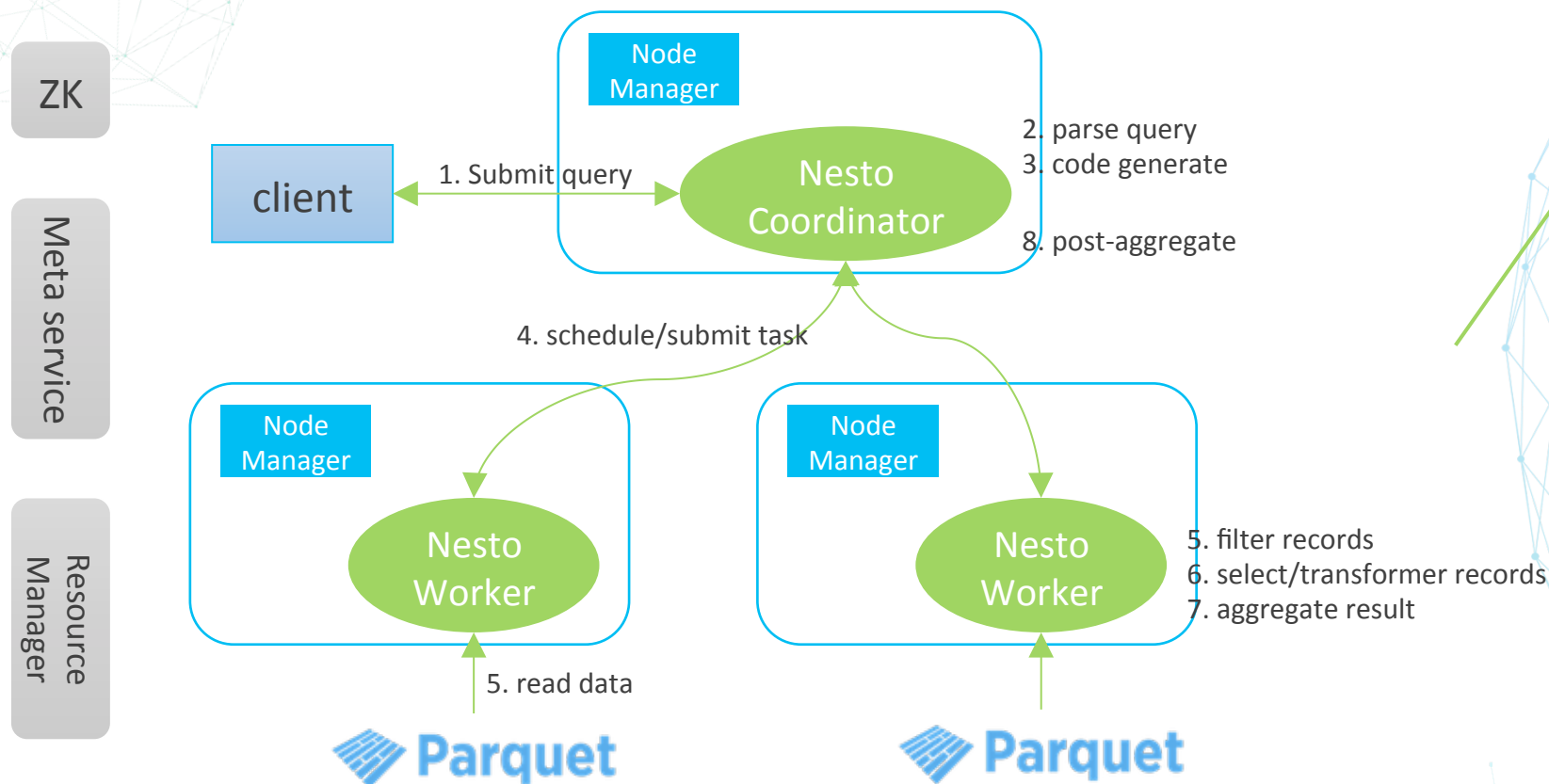
```
{  
  "userid": "cbjwilr98dkxk3lc88u3",  
  "int_attr1": "1998-08-08",  
  "int_attr2": "mark",  
  "int_attr3": 1,  
  "int_attr4": "paid",  
  "ad_exposure": [  
    {"content": 42212, "ts": "2016-11-15 08:25:11",  
     "campaign": 89, "creative": 3876}  
  ],  
  "watch": [  
    {"duration": 369, "platform": "xbox_one",  
     "ts": "2016-09-15 18:25:11", "show": 53}  
  ],  
  "label_churn_prob": 0.23,  
  "bk_tag1": true  
}
```

- User id as key

- nested data
- all historical data

- 3<sup>rd</sup> party and label data included

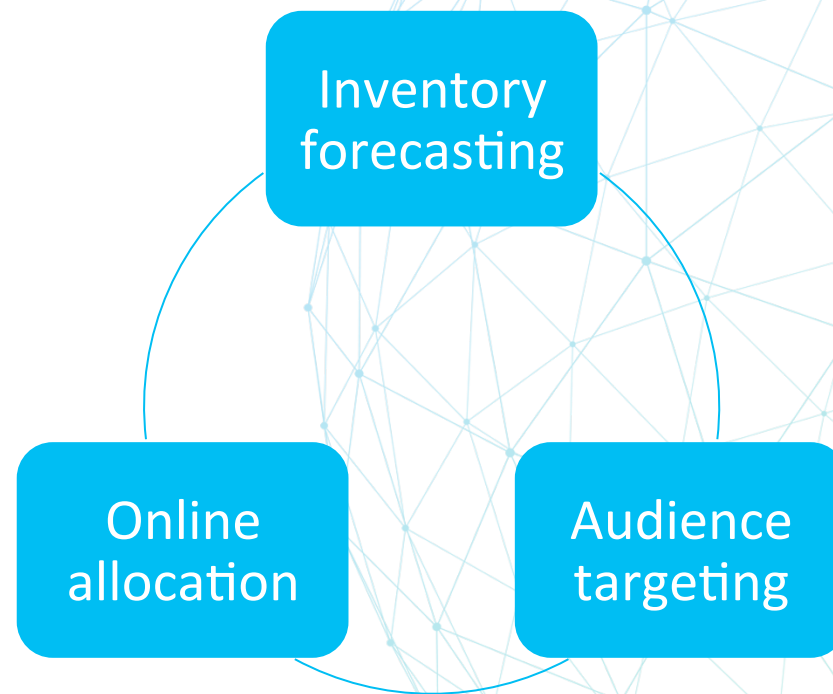
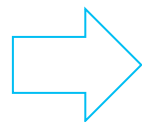
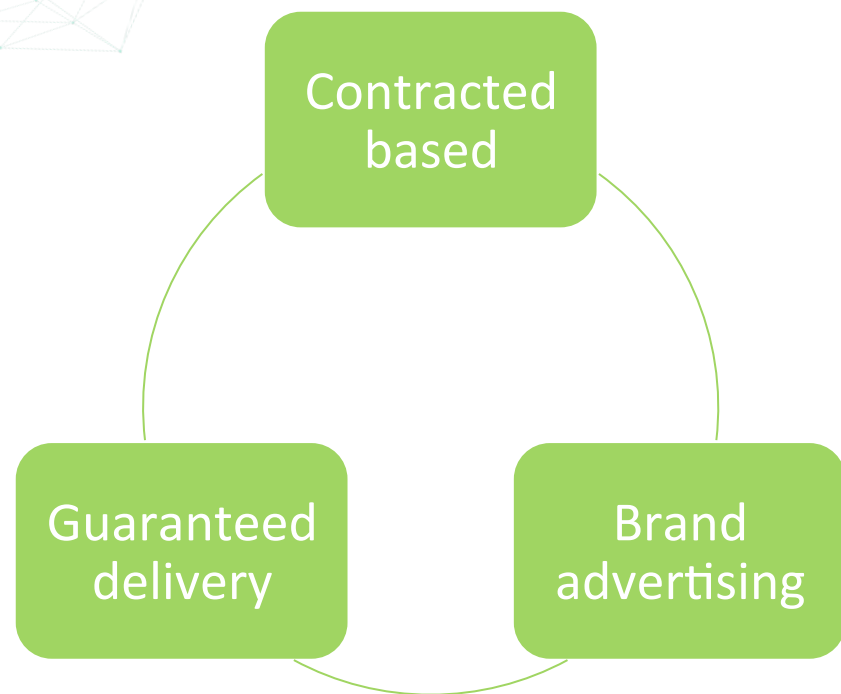
# Audience Platform – Analytics Engine (3)



1. Dedicated query engine for nested data
2. Optimized for advanced analytics

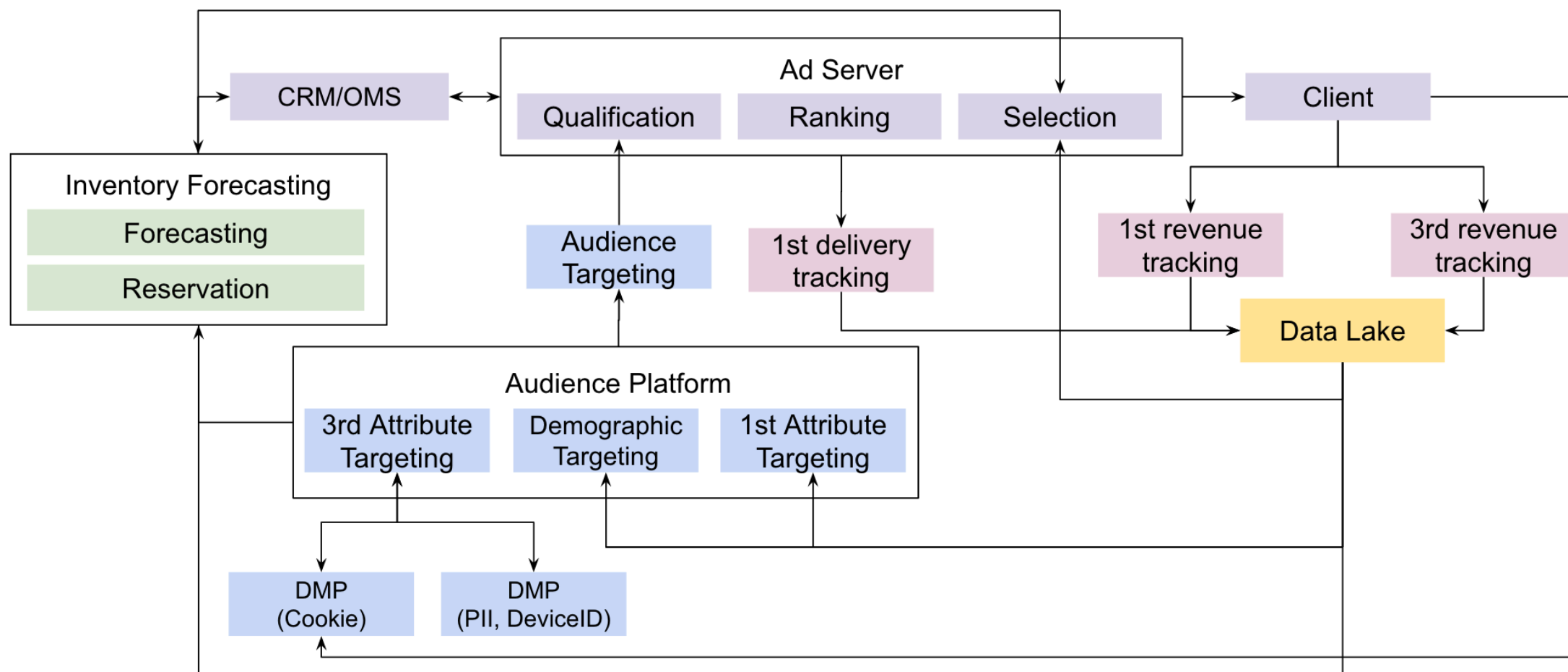
# Advertising In Hulu

# Advertising in Hulu – Overview

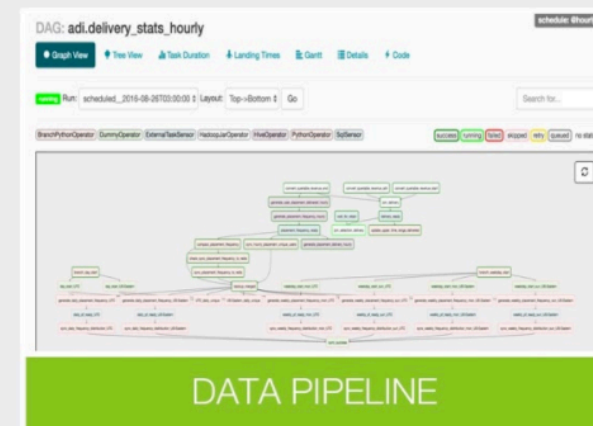
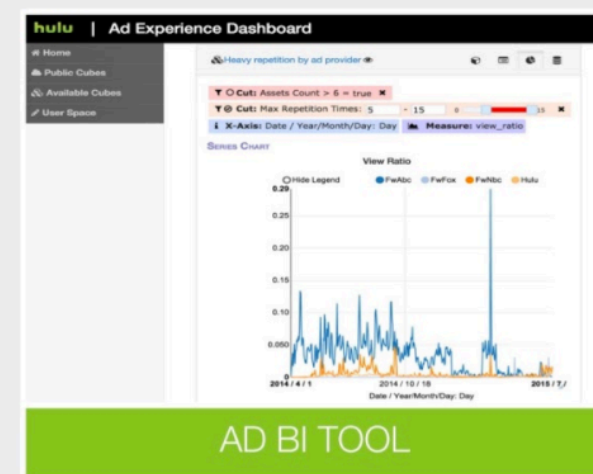
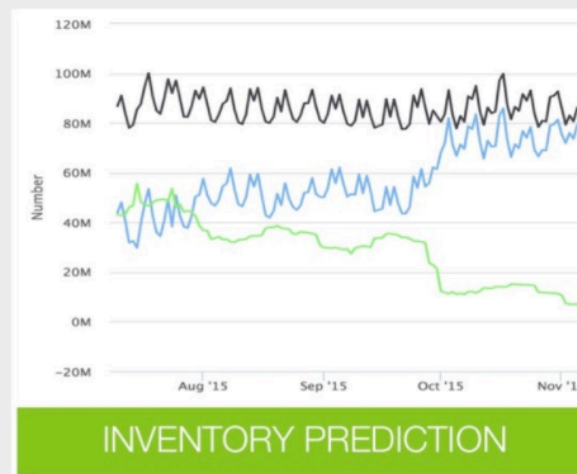




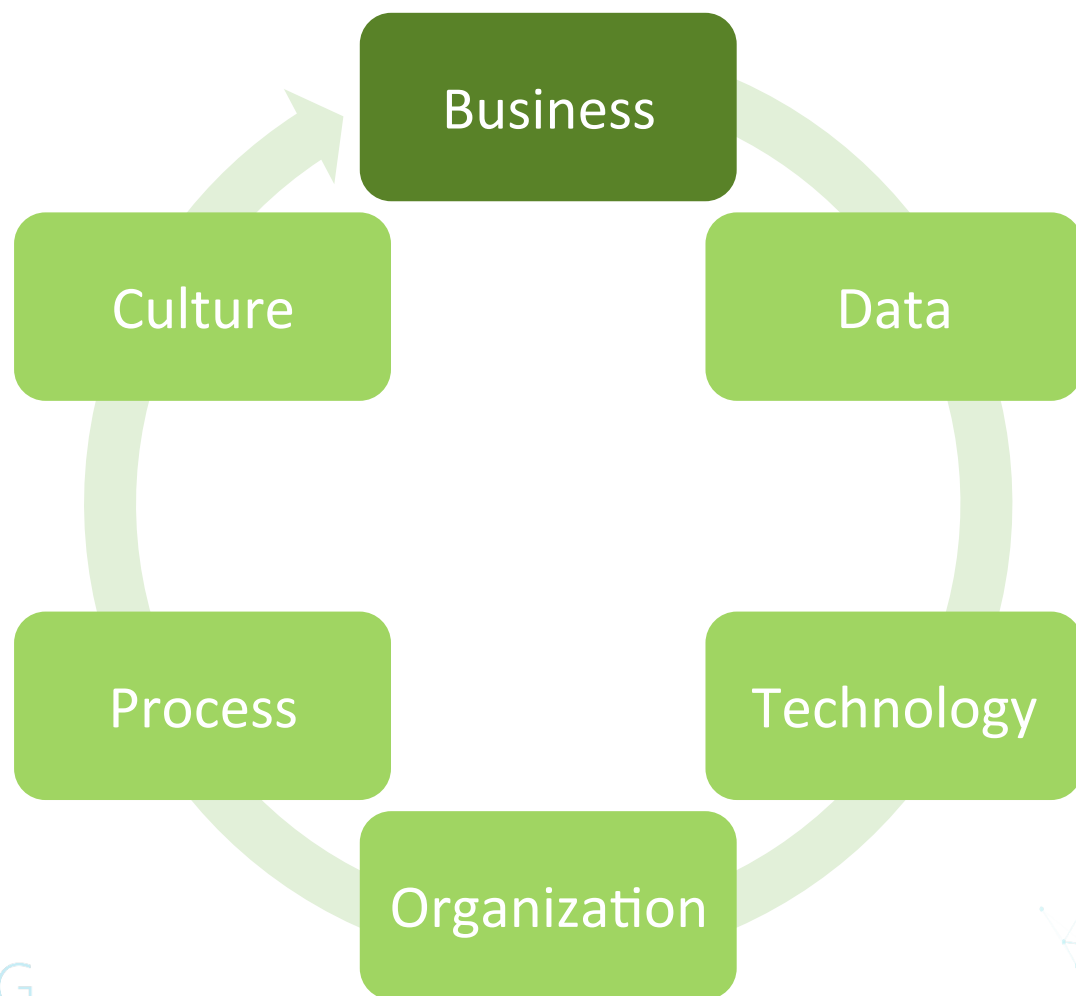
# Advertising In Hulu - Architecture



# Advertising In Hulu - Projects



# Summary



# Q & A



GIAC | BEIJING  
Dec.12.16-17

技术架构未来

架構  
ARCHNOTES  
高 可 用 架 构



Thank you!

