Quick Growth through ML Model A/B Testing

Introduce eBay **Experimentation Platform** for the Paid Search Ads

- Sleven Liu, Martin Zhang, Yi Liu

Agenda

- Why Growth hacking and A/B testing?
- Search Ads: The most important marketing channel
- Challenges and Solution for A/B testing
- Machine Learning Models Integration

Quick Growth in the eBay Paid Marketing through A/B Testing & ML Model

5+
Years

60+
Experiments/
Year

50+
Models/Year

Growth Hacking

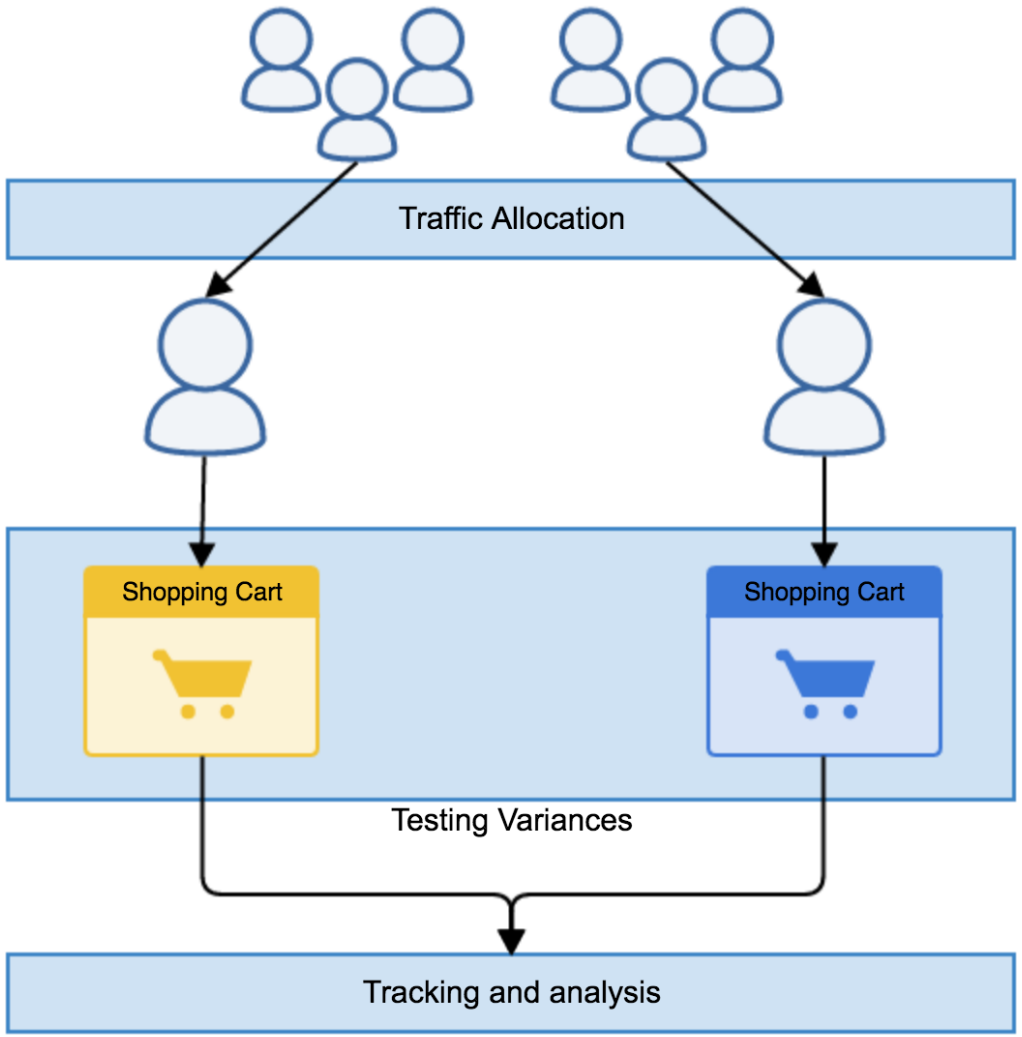
"Growth hackers are a hybrid of marketer and coder, one who…answers with A/B tests, landing pages, viral factor, email deliverability, and Open Graph.

On top of this, they layer the discipline of direct marketing, with its emphasis on quantitative measurement, scenario modeling via spreadsheets, and a lot of database queries."

- 《Growth Hacker is the new VP Marketing》
Andrew Chen

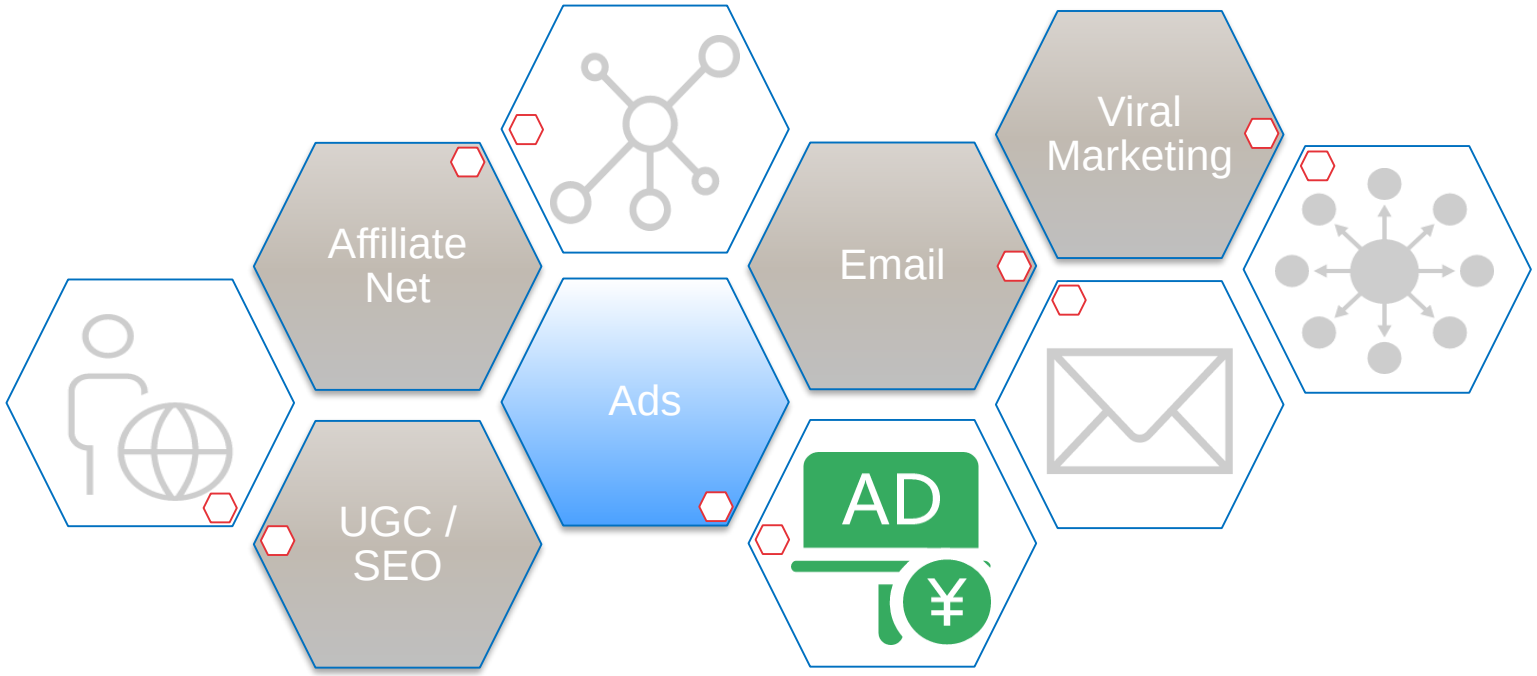
A/B Testing

- Key Elements
  - Statistical hypothesis
  - Sampling
- Benefits
  - Customer vs. expertise
  - Early launch and adoption in the marketing
  - Continue delivery and integration
  - Based on the data and statistics
- Limitation
  - Statistician Power
  - Imbalancing



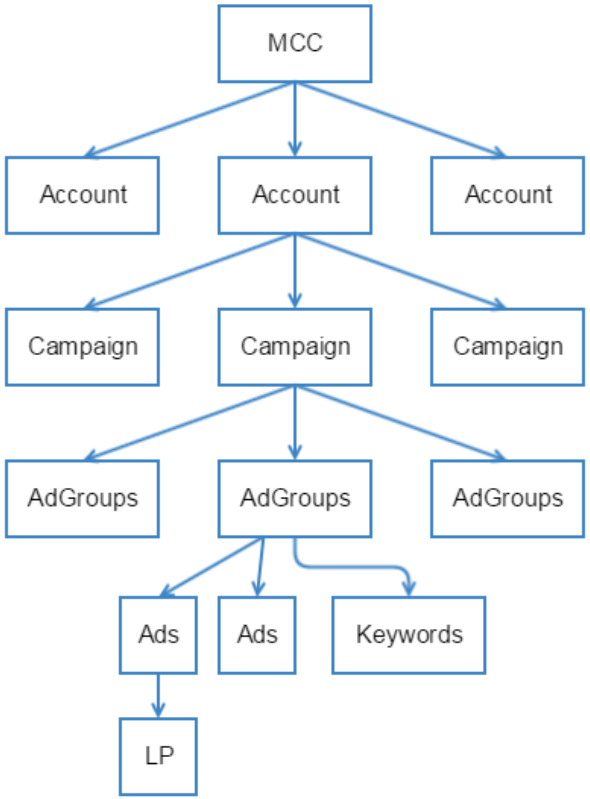
Growth Hacking Channels

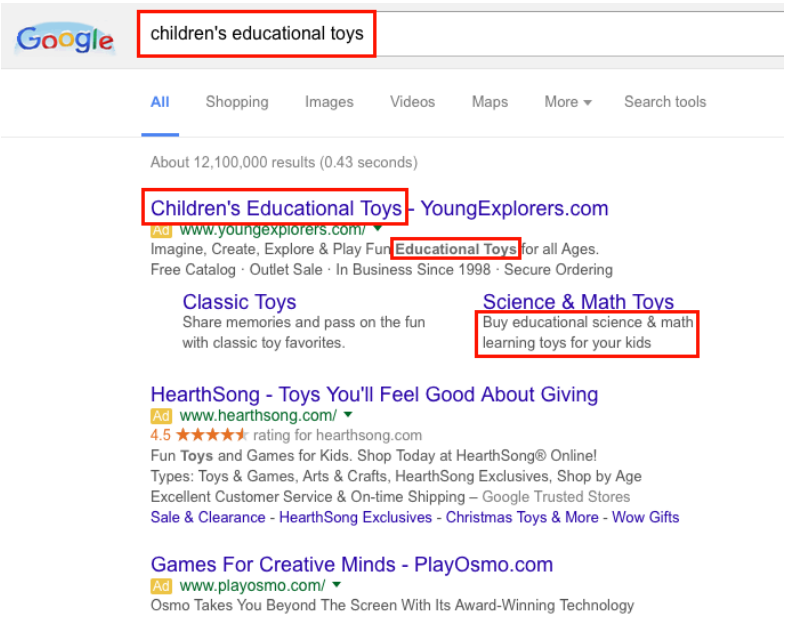
- “Poor distribution, not product is the number one cause of failure” – Peter Thiel, 《Zero to One》



Google Text Ads

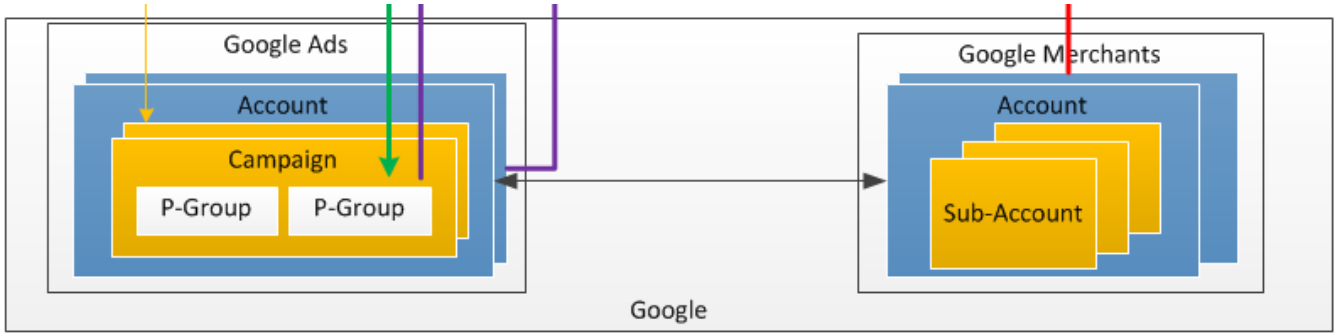
- Google Ads, CPC
- Content
  - Headline
  - Display URL
  - Description
- SRP + Search Network
- Exact vs. Broad match
- Campaign Structure

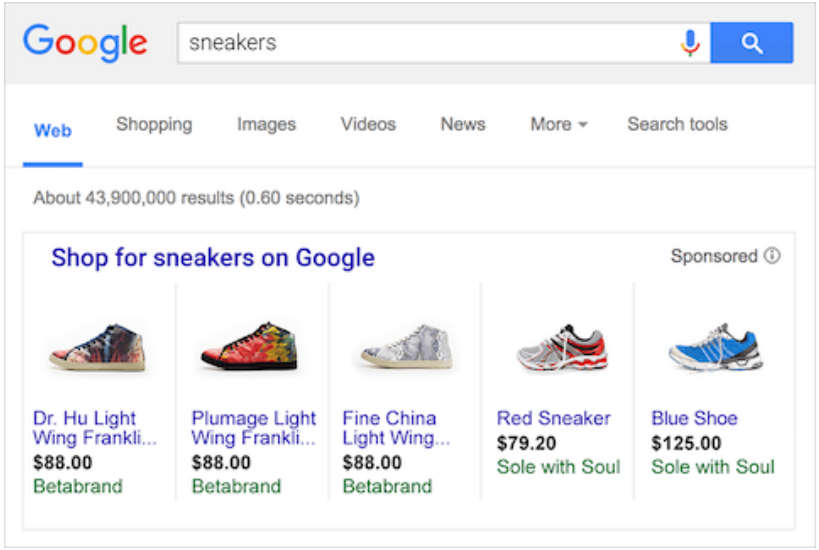




Google Product Listing Ads / Shopping Campaign

- More info (price/picture) more qualified traffic
- Catch more eyeballs
- Product/Brand match
- Higher barrier, less competition
- Backend structure





Challenges of A/B testing in the Paid Search Ads

Sampling

- No control on the user/visiting
- Accurate user targeting
- Skew data & Low coverage

Test Setup

- “Black Box” on third partner / ads platform
- Limitation of Testing objects

Tracking

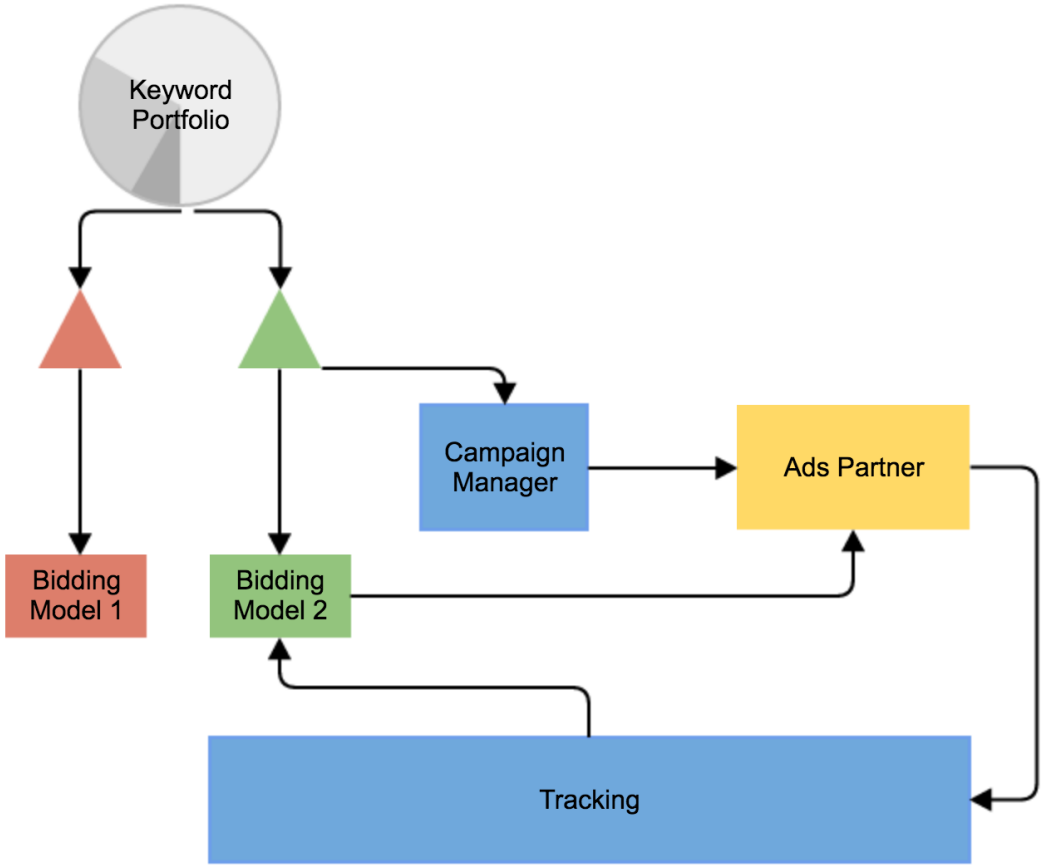
- External data loop

A/B Testing Solution Example in the Text Ads

- Based on the keywords
- Stratified sampling to resolve skewed data

- Campaign structure management
- Test object: bidding models

- Insides + outsides tracking
- Data loop for the model



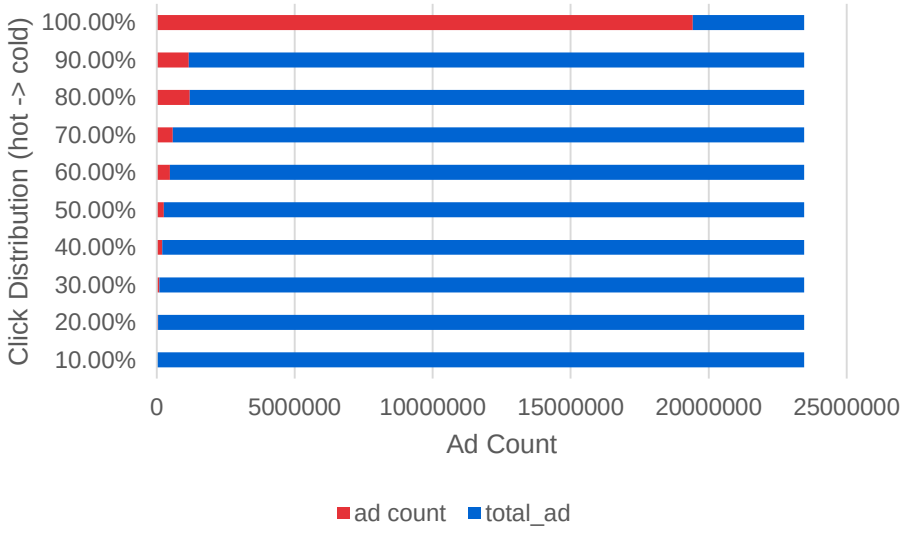
Why Sampling is important for A/B testing?

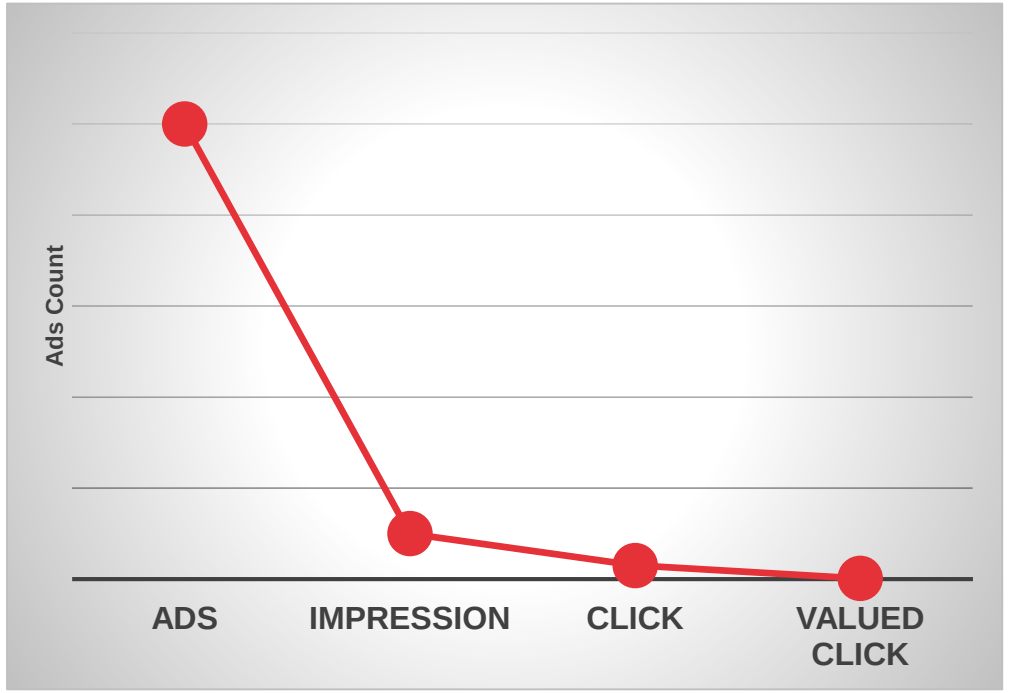
- Choose the right sample size
  - Is a large sample always good to speed up A/B? Or put business in real risk?
- Choose the right method
  - Why not using random sampling anyway?
- Un-represented sampling result might hurt business after rollout
  - Is the model workable for all the Ads? Or only the sampled ads?
- A trustable sampling result makes the A/B result trustable
  - Is the difference from A/B test result really from the model? Or because of the sampling difference?

Sampling Challenge – Huge volume of data

- Billion level Ads
- New Ads sourcing – is the process scalable for more ads added to marketing?
- Ads history tracking – how the process dealing with the historical data?

Sampling challenge – Skew Data & Low Coverage





- Top click queries
- Long tail queries

- Low Conversion Rate – Impression -> Click -> Transaction
- Deal with ads with no impression on partner

Sampling Solution - Method

Data Collecting — What data we can use?

Bias Filtering — Which keyword we can test?

Strata Building — How to evaluate the keywords?

Data Clustering — Which keywords are similar?

Sampling — Is keyword represented?

Validation — Is sampling balanced between A / B?

Sampling Solution - Tech

- Hbase + HDFS
  - Active ads stored in Hbase
  - Ads history stored in HDFS
- Spark
  - Huge data pre-aggregation
  - Optimization of huge data join with ads history, user behavior…
  - Store data as Parquet to improve the spark job efficiency

Machine Learning Model Integration

Where is the data?

What is a model?

How to manage the model lifecycle?

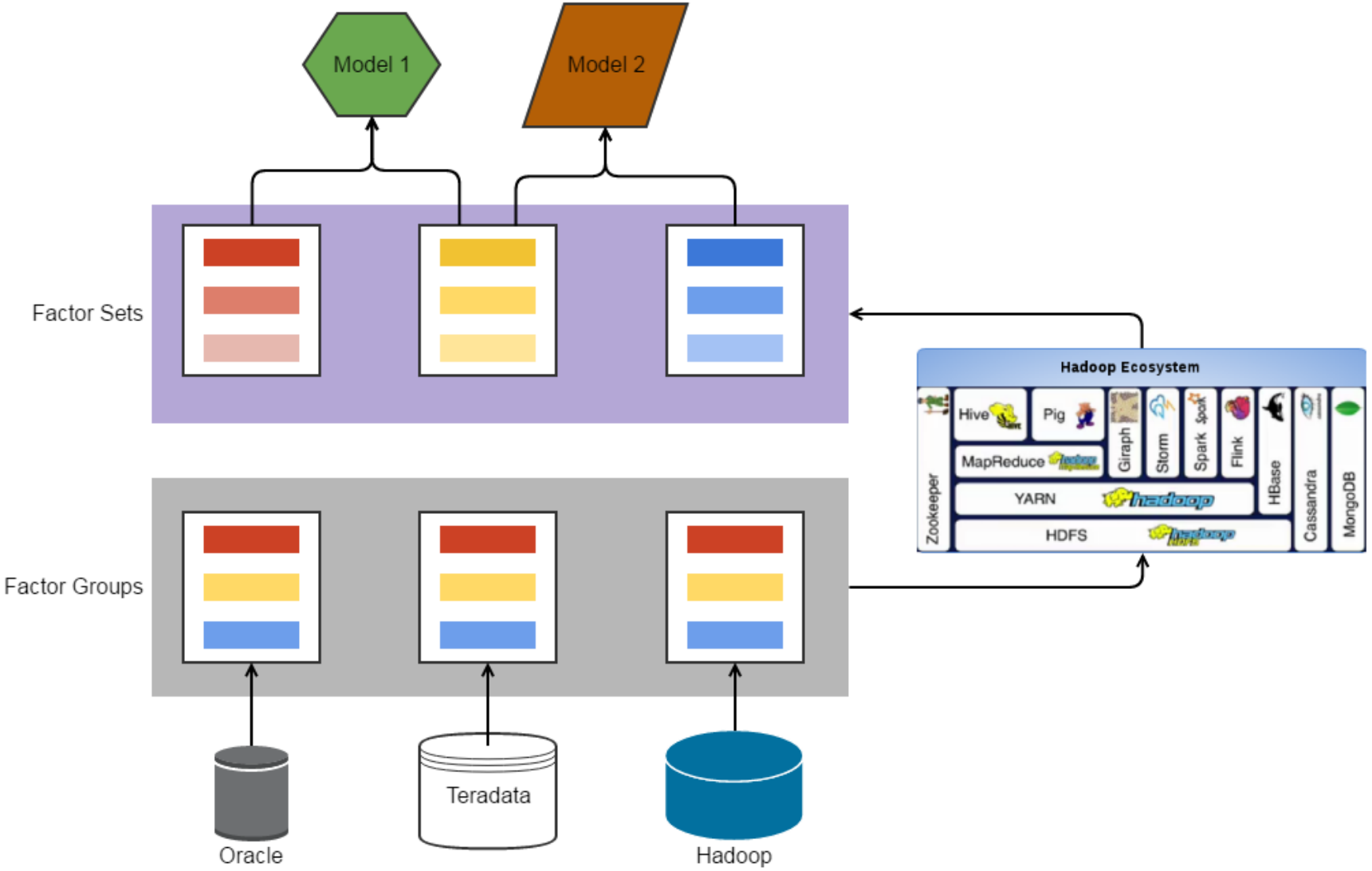
Challenge for data

- Data extraction
- Data processing
- Data gathering

- Original Solution
  - Regular ETL data pipeline to build factor for each model
  - Move gathered factors to model running env based on different scenario
- Bottleneck
  - Some effort are duplicated among different models
  - Factor is not reusable as it is built to meet special model's requirement
  - More effort to maintain the factor as it could be from different sources and built for specified model

New Solution - Factor System

- Factor: the model input
- Heterogeneous data sources
- Syntax + Semantic layer
- Calculate on the Hadoop
- Factor life-cycle



What factor system provides

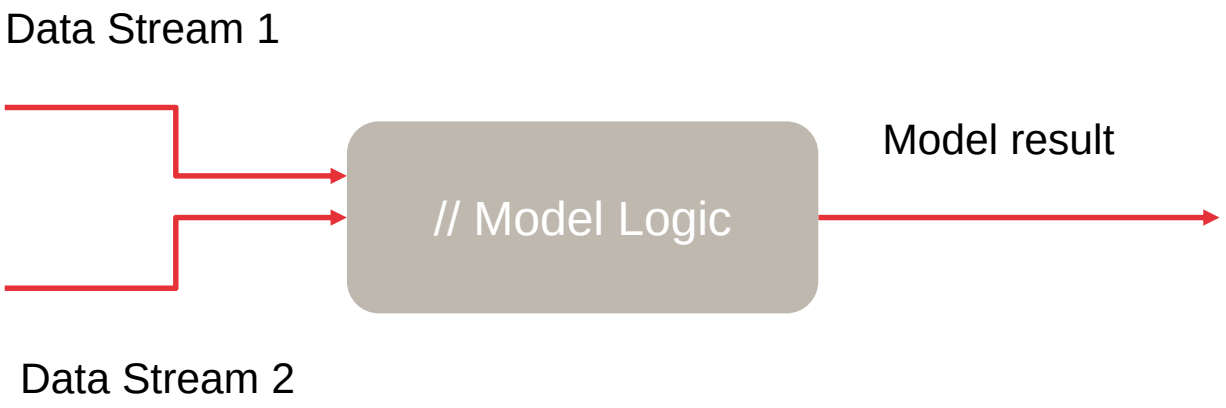
- Register Service
  - Factor code integration, deployment
  - External factor register
- Download Service
  - Online model input
  - Offline data exploring and model development
- Scheduling Service
  - Schedule the factor code in factor system due to different source data latency
- Dashboard
  - Factor status monitor, help understand the factor code running status
  - Factor meta definition, help data scientist better understand the factor to build the model

Capacity of Factor System

- PB level source data volume
- 10+TB daily increment
- 1000+ permanent factors, historical data backup on HDFS
- Use Cases
  - Batch Models - serve all the machine learning models for Paid IM marketing
  - Adhoc – to support offline data exploring for data scientist and data developer
  - NRT/Real-time (Future) - build factor cache for NRT or real-time model use cases

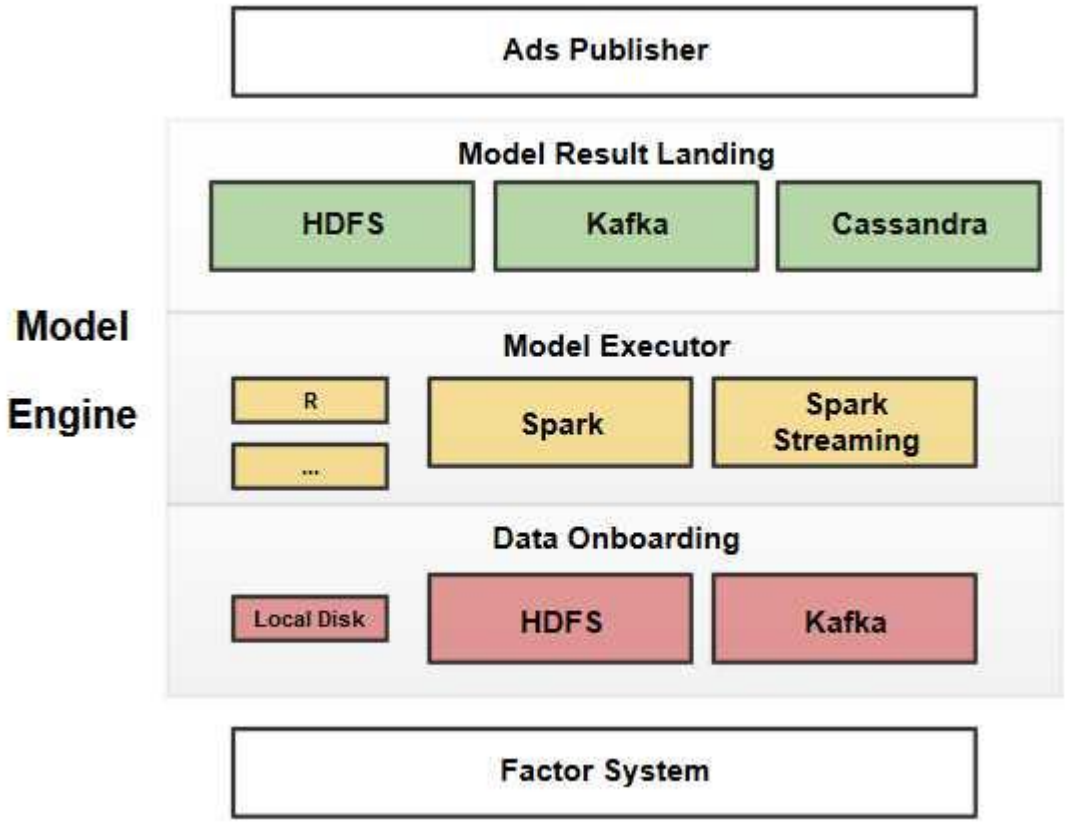
What model requires

- Model can access the wanted data based on the logical design
- Model can be executed in expected env using right tech to meet different use cases
- Model result can be delivered for real business needs



What is a model – Model Engine

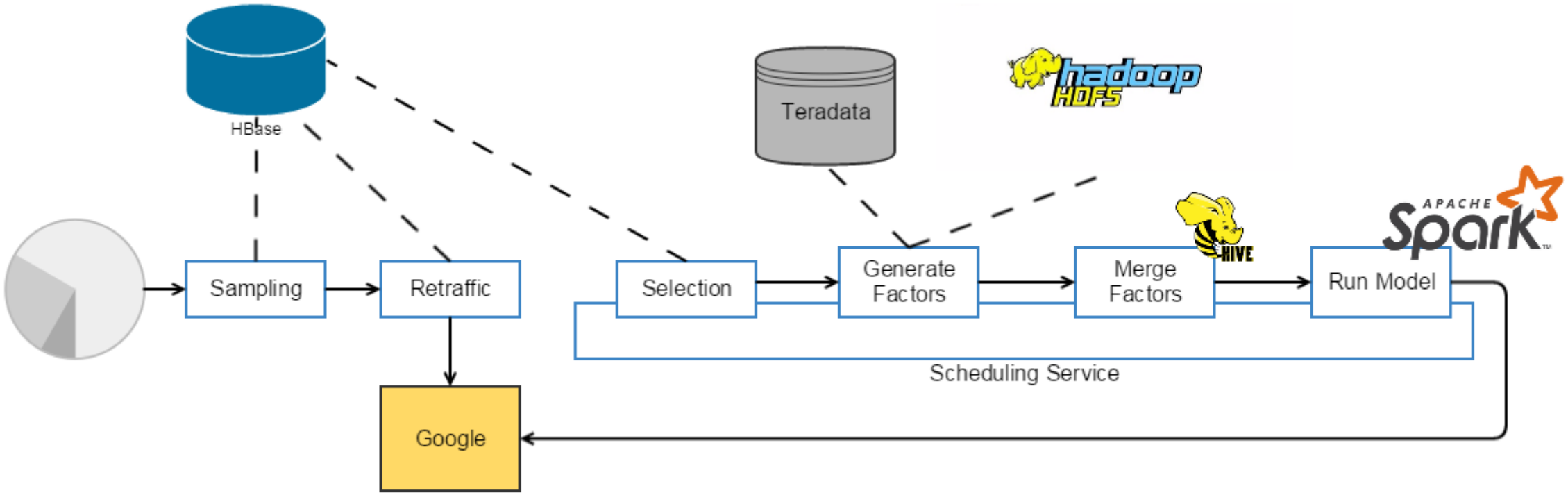
- Onboarding data from factor system to model engine
- Execute models using different tech solution to meet the real scenarios
- Landing result to different system to integrate with Ads publisher



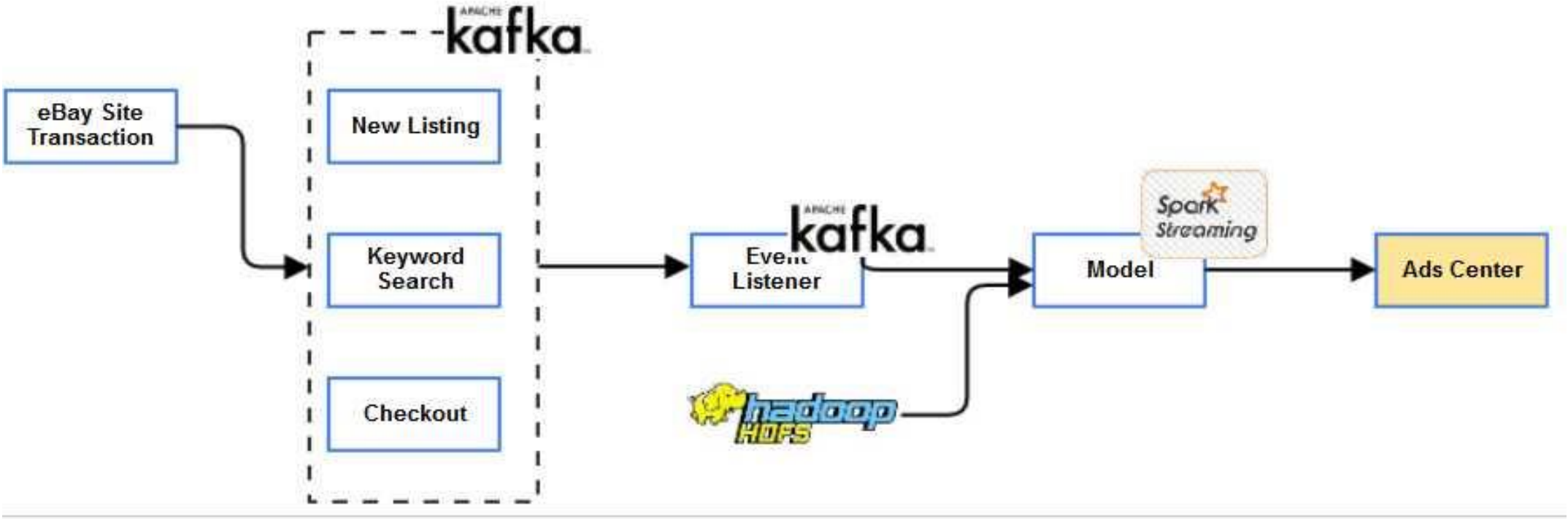
What model engine can help more to data scientist

- Sampled data for model training
  - Data scientist can get pre-sampled represented ads to train/test the models
- Real production factors access
  - Avoid duplicated effort from data scientist when developing new models with existing factors
- Self Service
  - Integration, provide staging environment similar to real-production for model execution to avoid integration issue after model deployment
  - Model deployment
  - Online debugging, all the model result/logs are kept in system to allow data scientist debugging during A/B testing
- Dashboard
  - Model status monitor

Model Lifecycle (Batch)



Model Lifecycle (NRT)



Anything Else for model?

- Is Model Result Reliable?
  - ➢ "SafeNet"
    - Collect the historical behavior of model
    - Detect any significant difference
    - Block the result sending to publisher

- How to track?
  - ➢ Ads Monitor & Alert
    - Expose online model result to Scientist/Analyst
    - Dashboard
    - Hourly & Daily report
    - Alerts deliver to model owner & business owner

Summary

- A/B Testing
  - Hbase, HDFS, MySQL, Oracle, Mongo
  - Java, Scala, SQL
- Machine learning model
  - HDFS, Kafka, Cassandra
  - Hive, Spark, Spark streaming
  - Java, Scala, R, Python
- Dashboard
  - InfluxDB
  - Grafana


Q
&
A