DSO459 Final Project

TIKTOK

Youyou Wang | Daniel Shen | Victor Ma | Leo Ma | Markenson Delkhaste

TABLE OF CONTENTS

DATA UTILIZATION

User

Content

Advertisement

Product

Audience Insight

Understanding user profiles based on users' personal information

Personalized Recommendation

Generate content recommendations based on user behavioral data

Precise Targeting

Utilizing user data to provide advertisers with accurate Ad placement services

Performance Analysis

Monitor the effect of Ads & provide optimization suggestions and data support for advertisers.

Content Optimization

Optimize content based on the video content itself and the data associated with it.

Content Management

Tagging & categorizing content, help improve the accuracy for content recommendations.

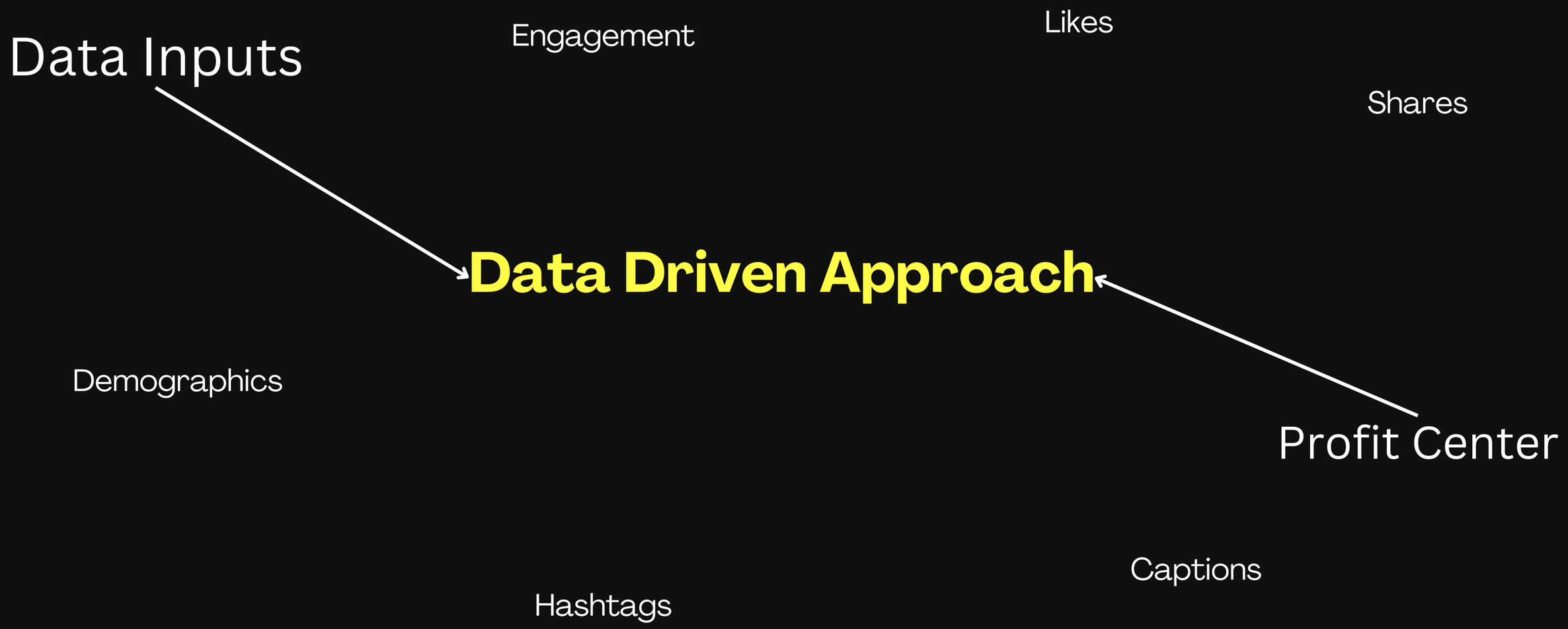
Product Development

Discover the pain points and users need to optimize the product interface, features and performance

Data-driven Decision Making

Support decision making for company's development.

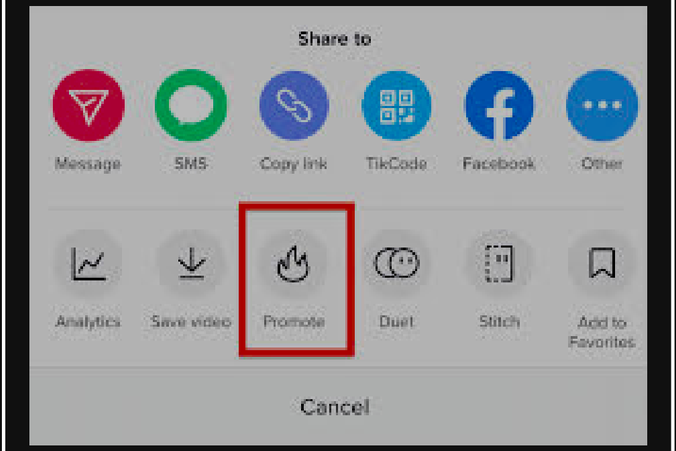
DATA UTILIZATION



BUSINESS MODEL

PAID PROMOTION (DOU+)

- Cost $10 - $7,000, one time
- Extended to potential audience



ADS PLACEMENT

Feed Ads

- Show up along with video, as link to advertiser
- CPM

Splash Ads

- Show up when users start the app
- cost over $10,000, depending on ad format, duration, launching time and area

COMMISSION

From Live Streaming

- Individual Streamers get 30%
- Guild streamers get 40 - 55%
- TikTok gets the rest.

From E-commerce

- Companies invite streamers to promote product sales
- TikTok gets about 10% of the sales

ANALYTICS in BUSINESS MODEL

- A conversion is any user action that has business value.

- Personalized experience

- “first-party measurement solution to enable web advertisers to go beyond the last-click model”
- Attribution Analytics

LIMITATIONS OF SPECIFIC STATIC MODELS FOR TIKTOK'S RECOMMENDATION SYSTEM

Qualification for a good recommendation system

Scalability

Flexibility

Non-stationarity

Able to learn from vast amounts of data

✕ SQL DATABASES

- Inefficient management of high-velocity user-generated data.
- Ineffectiveness in handling diverse, unstructured content like videos and comments.
- Challenges coping with the high volume of user data.

✕ LINEAR REGRESSION

- Inability to grasp complex, nonlinear interactions between user behaviors and content preferences.

✕ DESICION TRESS

- Increased complexity leads to a higher risk of overfitting.
- Difficulty accurately capturing dynamic & varied user preferences, potentially limiting prediction effectiveness.
- Compromising speed in making predictions.

RECOMMENDATION SYSTEM BEHIND TIKTOK

Monolith - TikTok's large-scale recommendation system

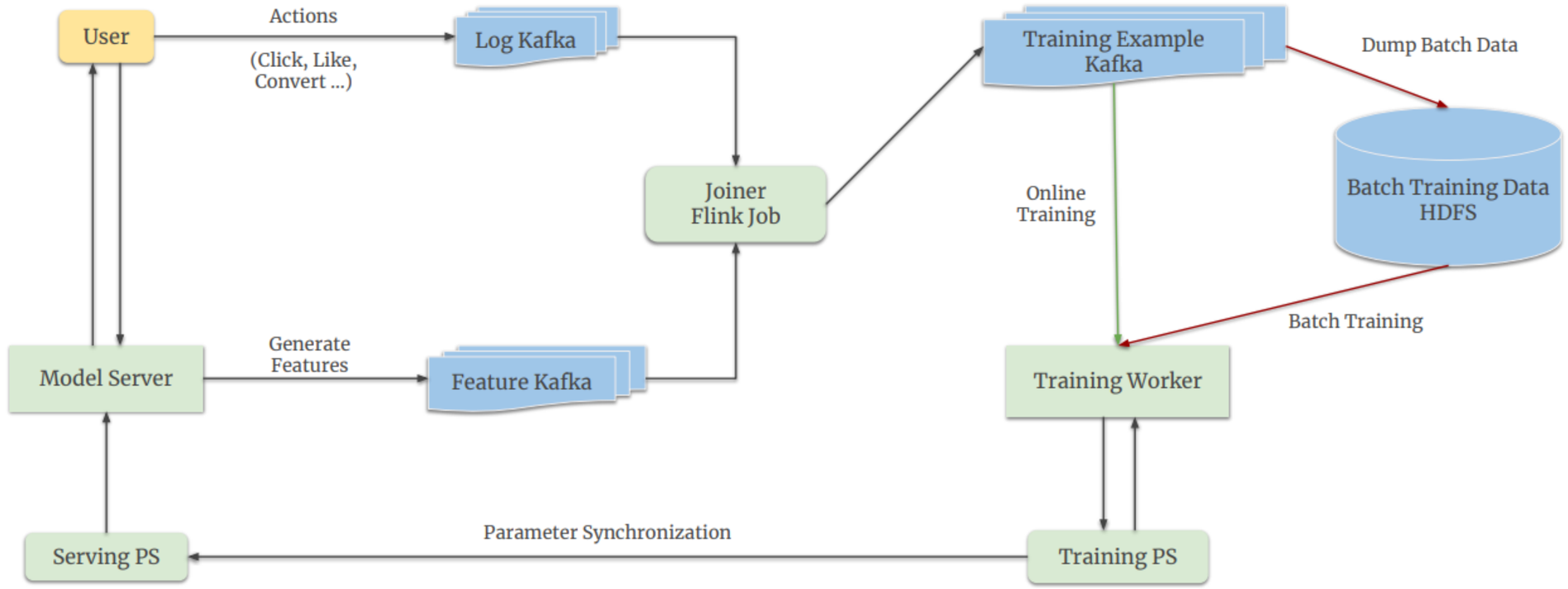


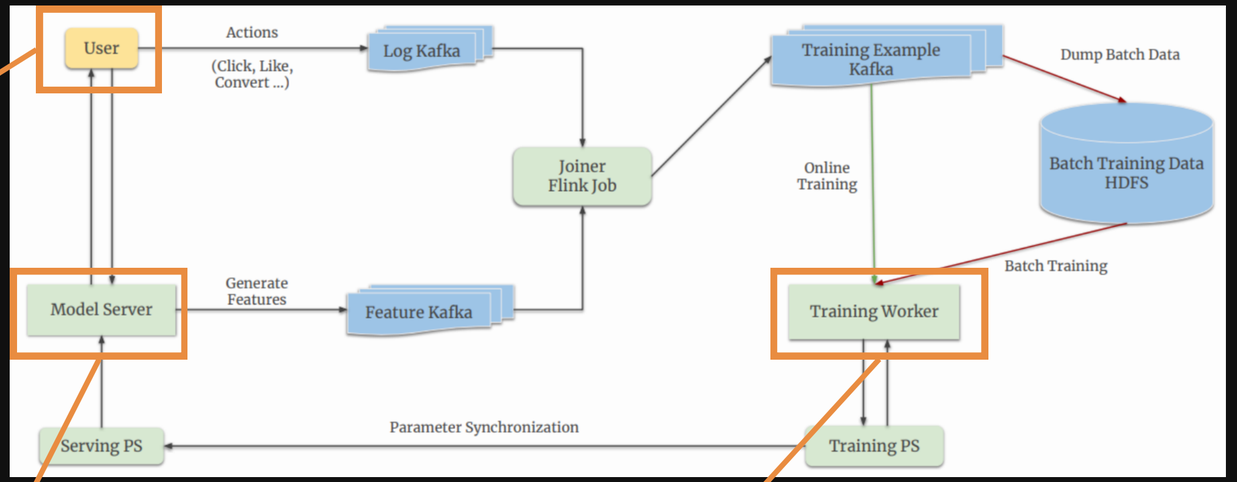
Figure 4: Streaming Engine.

The information feedback loop from [User → Model Server → Training Worker → Model Server → User] would spend a long time when taking the Batch Training path, while the Online Training will close the loop more instantly.

KEY PARTIES IN MONOLITH

USERS

- The start of the loop, where a real-world user interacts with a system.
- Users's interactions, such as comments, subscribe, or any other type of engagement, generate data.



MODEL SERVER

- Takes the user data and applies a predictive model to it.
- Uses features generated from the user data to make predictions and decisions, which include recommendations, personalizations, or other actions intended to improve the user experience.

TRANING WORKER

- Takes the user data and the output from the model server and uses it to update or train the machine learning model.

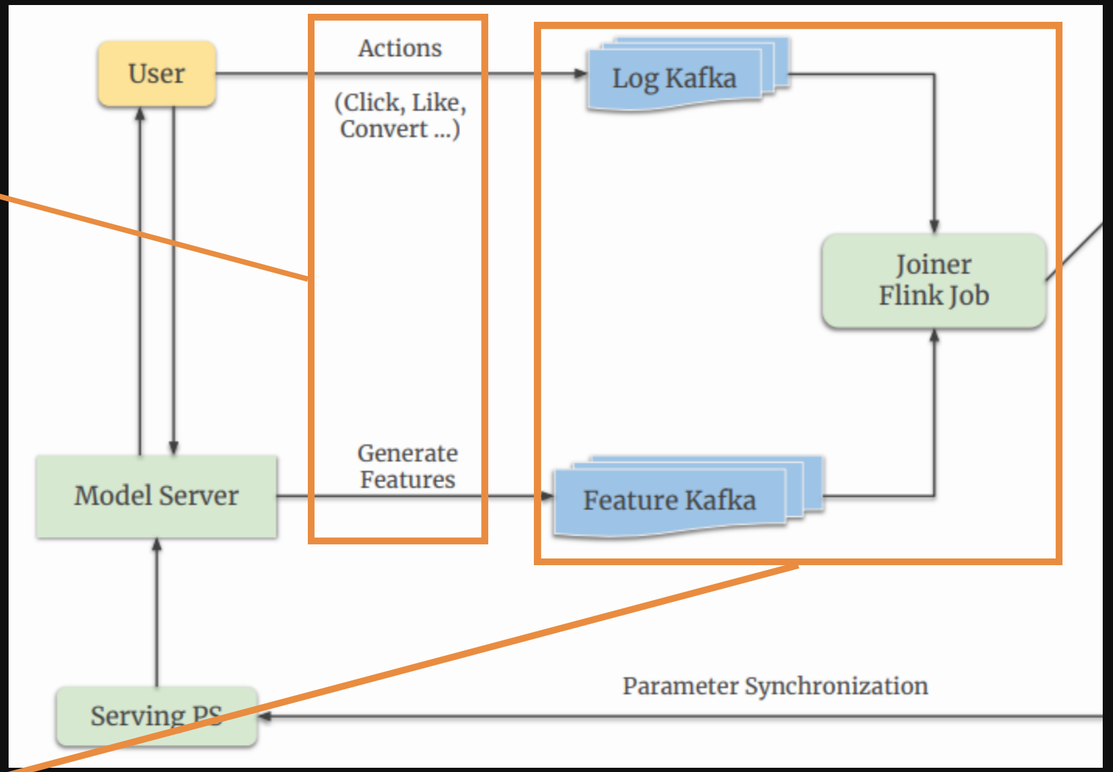
KEY ACTIONS IN MONOLITH

DATA COLLECTION

- User actions: Raw behavioral data such as likes, shares, comments, and uploads.
- Features: Structured and processed behavioral metrics derived from user actions such as peak activity times, frequency of shares, and duration of video views.

DATA INTEGRATION

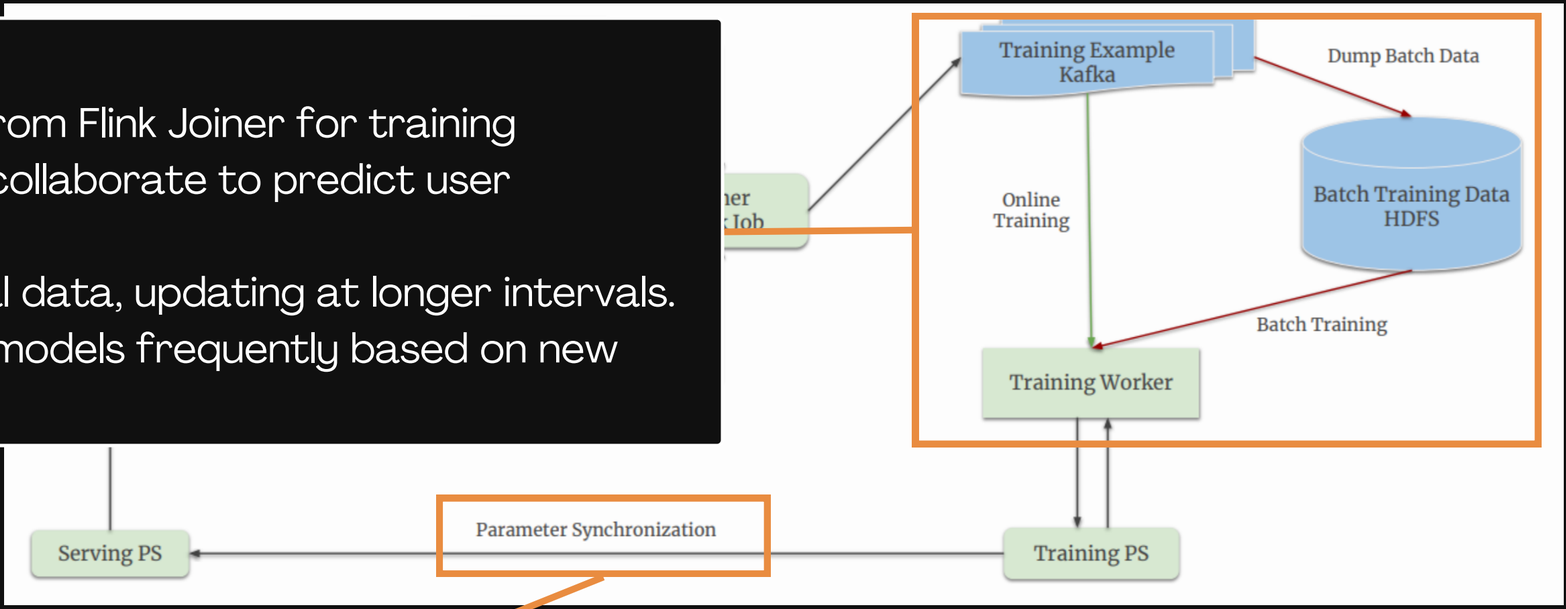
- Logs features and user actions separately in Kafka queues to handle high-volume data.
- Flink Joiner merges these logged actions and features together for processing.



KEY ACTIONS IN MONOLITH (CONT'D)

MODEL TRAINING

- Training Worker retrieves data from Flink Joiner for training purposes. Two types of training collaborate to predict user preferences:
- Batch training: Relies on historical data, updating at longer intervals.
- Online training: Updates existing models frequently based on new user data batches.



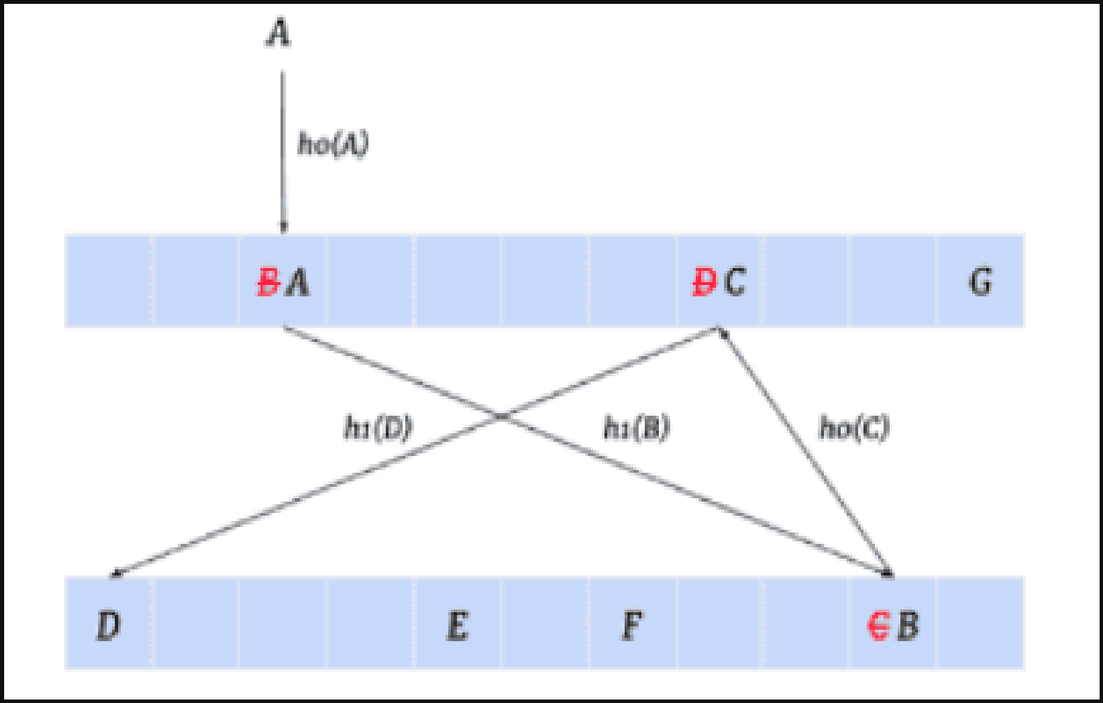
PARAMETER SYNCHRONIZATION

- Parameter Weighting: Determines the influence of each input on predictions (e.g., the weight of user comment frequency on watched videos).
- Model Server Parameters are updated iteratively under the Training Worker's guidance. After comparing predictions with actual outcomes, updates are made.
- Fundamental parameters are updated less frequently to reduce the model workload while maintaining efficiency.

COLLISIONLESS HASHING IN TIKTOK

Collisionless Hashing:

- ***a method of organizing data where each piece of information is assigned a unique 'shortcut' or hash, avoiding overlaps or 'collisions'.***
- ***This makes data retrieval and management more efficient, especially in large-scale systems.***



Application in TikTok:

- Managing User Data: TikTok handles vast user interaction data (likes, comments, views) efficiently through collisionless hashing techniques.
- Cuckoo Hashing: TikTok likely utilizes cuckoo hashing with dual tables and hash functions. If a collision happens in one table, the system tries placing the data in the other until unique placements are achieved.
- Enhanced Recommendation System: Collisionless hashing enables TikTok to swiftly analyze user interactions, enhancing accuracy in video recommendations for a better user experience on the platform.

APPENDIX

JOB TITLES & DESCRIPTION

Job Title	**Job Description**
Data Engineer	• Design and build data transformations efficiently and reliably for different purposes (e.g. reporting, growth analysis, multi-dimensional analysis); • Design and implement reliable, scalable, robust and extensible big data systems that support core products and business; • Establish solid design and best engineering practice for engineers as well as non-technical people.
Data Scientist	- Drive insightful understanding of TikTok user and creator ecosystem. - Design and analyze experiments to verify product hypothesis and direction. - Audit quality of launch experiment results and guarantee correct measurement of product or technical changes' impact. - Keep track of core metrics and attribute metrics changes to root causes. - Use data and experiments to verify hypotheses around bottlenecks in product growth, and consult on the direction for improvement. - Work with engineers and product stakeholders to deliver product insights and strategy.

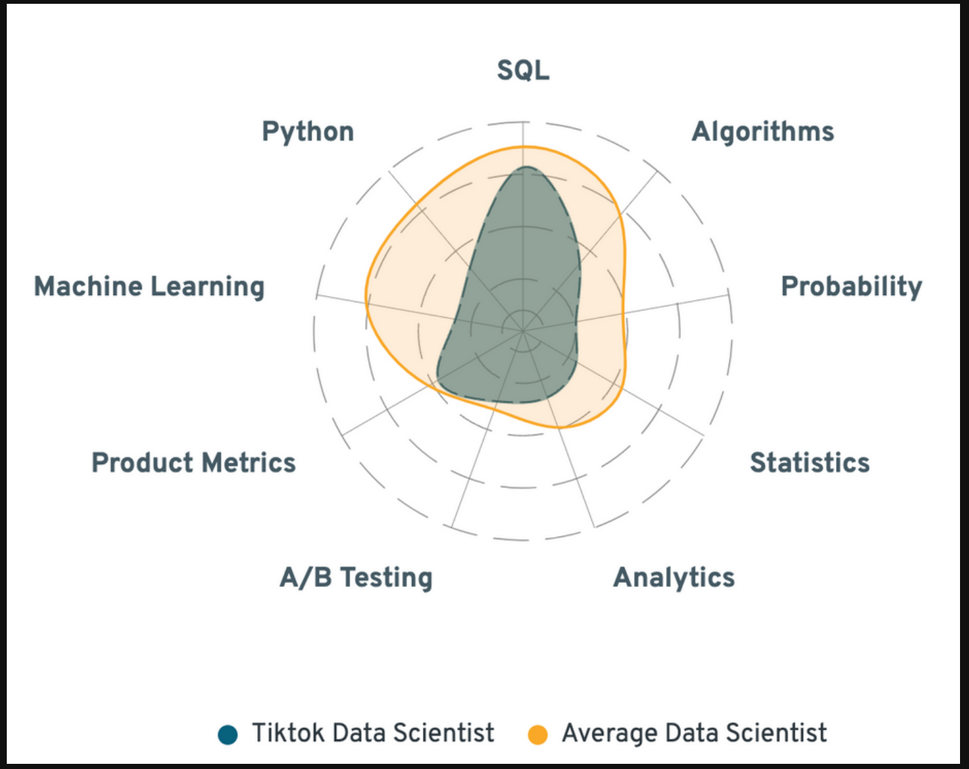
AVERAGE JOB QUALIFICATIONS

1. Bachelor's degree in related technical field, or equivalent practical experience

2. Experience in coding languages, including Python, Java, C++, Swift, Objective C, JavaScript, etc

3. Strong understanding of algorithms or good communication serve as a plus

4. Previous experience in an internship/research/project in that particular field is preferred



EXPECTATIONS

WORK-LIFE TRADEOFF

All work, relentless pursuit of productivity, balance nonexistent

SALARY EXPECTATIONS

Median salary $205,000, however even for lower level/entry level position you can expect six figures

BASE VERSUS BONUS

Year-end bonuses besides typical six figure+ salary

TIME OFF OR FAMILY LEAVE

10 paid holidays/year
17 days of Paid Personal Time Off (PPTO)
10 paid sick days/year
12 weeks of paid parental leave
8 weeks of paid Supplemental Disability.

RETIREMENT PACKAGES

401k matches 5% up to your 10%, offers stock options

COMPANY CULTURE

High turn-over rate, burnout is common
Largely influenced by parent company ByteDance, which creates a much more chinese style culture emphasizing work

MEDICAL INSURANCE BENEFITS

100% premium coverage for employee medical insurance, approximately 75% premium coverage for dependents and offer a Health Savings Account(HSA) with a company match.
Dental, Vision, Short/Long term Disability, Basic Life, Voluntary Life and AD&D insurance plans, Flexible Spending Account(FSA) Options like Health Care, Limited Purpose and Dependent Care.

EXPECTATIONS

ByteDance Reviews

3.1 ★★★☆☆ 1,196 reviews

5 ★
4 ★
3 ★
2 ★
1 ★

Ratings by Category

3.2 ★ Career Growth
2.5 ★ Work - Life Balance
3.4 ★ Compensation / Benefits
2.7 ★ Company Culture
2.7 ★ Management

4.0 ★★★★☆ “Late nights and lots of broken things ”

Current Employee · n******** · Project Manager - Oct 23, 2023

1.0 ★☆☆☆☆ “Management is very toxic and harsh ”

Current Employee · n***** · Project Manager - Oct 23, 2023

2.0 ★★☆☆☆ “Good for new grads”

Current Employee · b****** · Engineer - Oct 20, 2023

INTERVIEW

PROCESS

Recruiter screen
> tech screen
> tech interview
> tech interview
> recruiter screen

RATING

EXPERIENCE

"I declined to move forward after this round because they took almost 2 months to get back to me, and after 2 months they wanted me to progress to the next round. ***Was too long of a wait****."*

"The really negative experience came when dealing with the HR team. They were really aggressive about asking what my current comp is, even going so far as to insist I sent them my current compensation letter (illegal in my state, and most states). They also wouldn't give me a compensation range for the level / role (legally required to do so in California). They were also really secretive about role levels, and couldn't share anything about what levels meant, or what it takes to get to the next level."

OVERALL TAKEAWAY

After this research process, we found that:
We are willing to go further into the data analytics field, but not with TikTok.

WHY THIS FIELD:

Interest in data and technology:
Constantly evolving field with opportunities to learn new techniques, tools, and technologies.

Passion for Innovation:
Opportunities to innovate through data-driven decision-making and discovering new trends.

High Demand & Competitive Salary

WHY NOT TIKTOK

High Risk & Low Stability
The data privacy issue is still unsolved with the U.S Gorvenment. As a foreign company, this is risky

Low Work-Life Balance & Poor Company Culture
High potential of deploiting the employees. Reviews online showed no obeyance to the labor law