

HAIL, HYBRID!

**A BETTER
RECOMMENDATIONS
ENGINE**

May 2019

WHAT WE'LL COVER TODAY

- Who is SoftServe
- Why We Started This Journey
- The Modern Hybrid Recommendations Engine
- Data Science Behind It
- What We Learned

ABOUT SOFTSERVE

8,000+

EMPLOYEES

We hire 10-12 daily

30+

OFFICES

Europe & USA

12

COUNTRIES

With SoftServe Clients

25

YEARS BUILDING

ENTERPRISE SOFTWARE

Founded in 1993

4,500+

LARGE SCALE
CUSTOMER
PROJECTS

6,000+

SOFTWARE ENGINEERS

700+

CLOUD
EXPERTS

300+ Cloud DevOps Focused

70

NET PROMOTER
SCORE (NPS)

More than twice as high as
our competition

softserve

KEY TECHNOLOGIES AND PRACTICE AREAS

experience design

internet of things

big data,
artificial intelligence,
and machine learning

cybersecurity

cloud/devops

digital experience platforms

mixed reality

blockchain

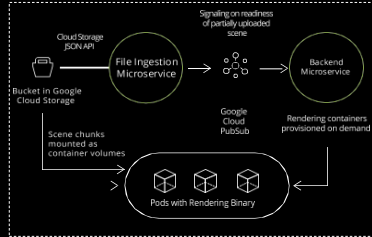
robotics

softserve

PROBLEMS WE SOLVE IN MEDIA

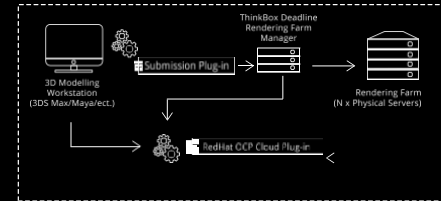
Video Content Supply Chain

Capture and Ingest
Analyze and Manage
Transcode and Distribute



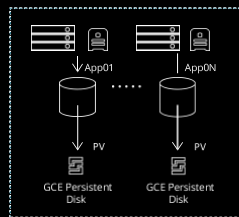
Cloud Based Workflows

Rendering Farms
Transcoding
MAM / DAM Solutions



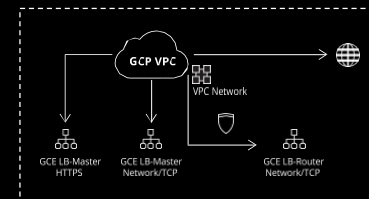
Data Analytics

Data Workflow / Data Pipeline Design
Business Solutions (recommendations, churn prediction, attribution)
Machine Learning & AI



DevOps and Automation

Serverless Architecture
Cloud Native and Containerization



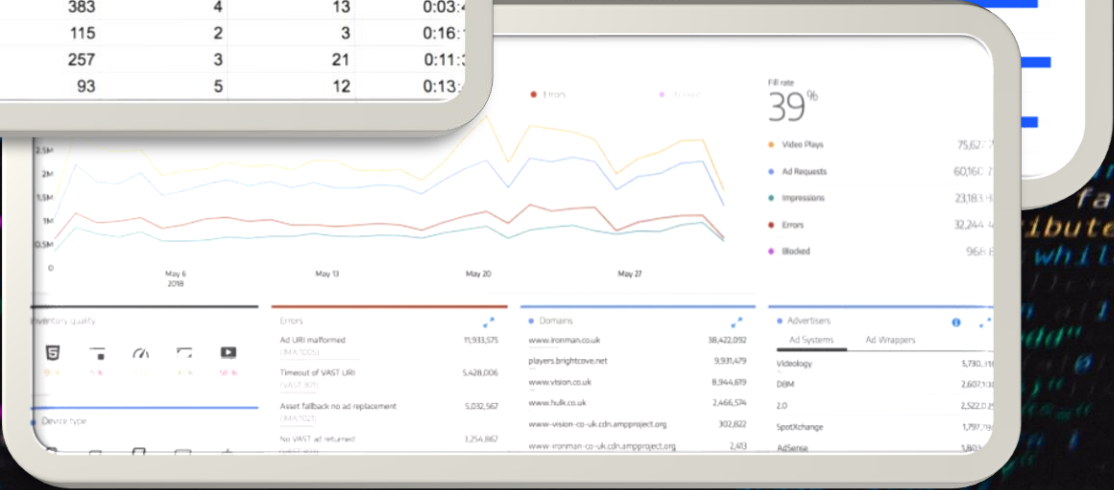
softserve

WHY WE EMBARKED ON THIS JOURNEY

Field	Type	Description	Value
date	timestamp	Date and time the original file was written to S3	2018-01-01T00:00:00.000+0000
event_date	timestamp	Date and time of watching the event	2018-01-01T00:02:24.000+0000
year	string	Year of the event	2018
month	string	Month of the event	1
day	string	Day of the event	1
seriesid	string	Unique identifier provided by Gigya	f4bcb3b86cb9425c9d3e62798f5ee8e2
site	string	Defines which platform the content was watched on	iosplayer
videoname	string	Name of programme	The Chase
episode	string	Name of episode	Mon 04 Dec, 5.00 pm
videomatid	string	Unique identifier for content, ITV & STV content has a different format	
videogenreid	string	Content Genre	
videolength	float	Clip Length/Runtime - This is the maximum duration or length of content being consumed (in seconds)	
streamid	string	Unique identifier for each stream	
broadcastdate	timestamp	Date and time the original programme was broadcast	
appversion	string	Define which version to the app the programme was watched on	
mobiledevice	string	Name of the actual device	
mobiledevicetype	string	Name of the generic type of device	
deviceid	string	Id associated to mobile device used - only relevant to android mobile devices	
visitorid	string	Adobe or Comscore unique identifier	
customcontenttype	string	Live, Shortform or VOD	
visitvideosequence	string	An incremental counter to allow analysing which video was first, etc.	
adname	string	If the content was an advert then when was it played, pre-roll, mid-roll, post-roll	
adid	string	Was the content an advert	
browser	string	Browser type and version the content was viewed on	
platform	string	Platform: web, mobile and connect tv	
adlength	float	Length of video ad in seconds	20
adloadevent	float	Load event for the video. (This occurs when the viewer clicks the Playbutton). Would count even if there are pre-roll ads, buffering, errors, and so on.	0
adstarttime	float	Duration of the video viewing AND video ads, minus buffering, and seeking aggregated viewing time, in seconds	3033
adstarttimestamp	integer	First frame of video is viewed. If viewer drops during ad, buffering, etc then there would be no "Content Start" event	1
adviewtime	float	Duration of the video viewing, minus the video ad, buffering, and seeking aggregated viewing time, in seconds	2727
adstarts	integer	Number of video ad starts	0
adcompletes	integer	Number of video ad completes	0
adviewtime	float	The total amount of time, in seconds, spent watching the ad	0

Short Title	ID	V	L	D	C	##	Duration
Intro	zO9U88i2S1M		7,706	269	6	14	0:00:4
HTTP Archive	uy1ZVNzo5KY		8,406	348	4	7	0:03:4
JS Vulnerabilities	llwgeZgZFB0		12,451	457	20	30	0:12:4
CrUX	-oAhPs3ooV8		7,957	241	8	7	0:04:2
WordPress Perf	_MyFUvB4mhE		5,034	100	37	24	0:03:8
FID	ULU-4-ApCjM		10,473	308	6	6	0:03:4
dev.to	p1JAIOsewWB		7,379	215	19	12	0:10:4
HTTPS	kBzbKUirOFk		11,030	365	16	26	0:08:8
WPT	LXfL2QyVrR4		7,965	383	4	13	0:03:4
Transparency	hqTtkdNwYwk		3,594	115	2	3	0:16:3
Web Fonts	L0xE0dybGFQ		6,903	257	3	21	0:11:3
DEV Survey	lx179QNRRvs		3,266	93	5	12	0:13:3

Action	Date	Name	Email	IP Address	Play Graph
			ben@pirate.org	198.36.213.26	
			eric@juicypcar.co.uk	70.182.111.139	
			steven@yahoo.com	104.219.198.135	
			glenn@greenshrub.com	75.72.230.236	
			zach@google.com	198.36.213.26	



DATA TYPES

A woman with blonde hair is sitting in bed, looking at a laptop. The room is dimly lit, with the laptop screen providing the main light source. The text is overlaid on the image in various colors and fonts.

Content Data

Player Data

Geographic Data

Conversion Data

softserve



**I AM AN
INDIVIDUAL**

NOT A DEMOGRAPHIC

INTEGRATED DATA TYPES

Trends

Social graphs

Linear past consumption habits

Interests in celebrities, topics, genre, etc...

Gender

Demographics

Time user is active

Device type(s)

Other interests



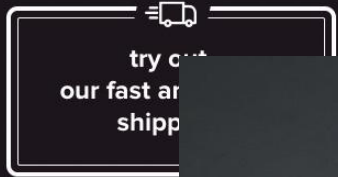
BRAND

it's all about style

new trendy line
autumn-winter 17/18

SHOP NOW

money back guarantee



BRAND

Spring-Summer 2018
Collection

Be Stylish. Be Yourself

Cashmere Sweater
\$35

SHOP NOW

money back guarantee

CONSUMERS PSYCHOMETRIC PROFILING

Innovative retail personalization approach that leverages **OCEAN** psychometric concept to help retailers identify customer personality profiles and suggest the best way to communicate and engage with the consumers.



Shopping Habits

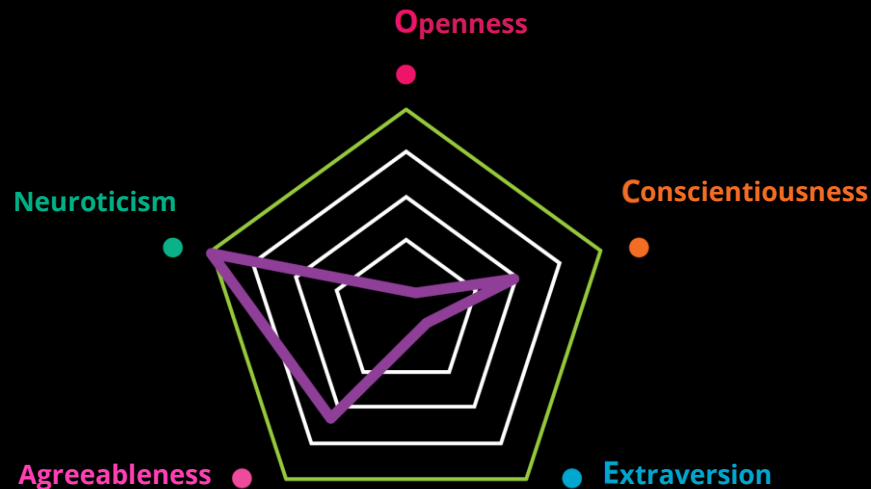


Verbatim Data
from Social Media

LEARN MORE

softserve

It combines data analytics with Ocean/BIG5 approach and technology to reach a deep understanding of the customer.



BIG FIVE PERSONALITY TRAITS

Openness

Extent to which person is open to experiencing a variety of activities

Conscientiousness

Person's tendency to act in an organized or thoughtful way

Agreeableness

Person's tendency to be compassionate and cooperative towards others

Neuroticism

Person's emotions are sensitive to the individual's environment

Extraversion

Person's tendency to seek stimulation in the company of others

softserve

WHAT PROBLEMS ARE WE SOLVING

Customers have more choices than ever to watch content

Keep them on your platform, and keep them engaged and coming back.

Create a competitive advantage



softserve

CREATING A COMPETITIVE ADVANTAGE

Revenue Driven

Customer Experience Driven

softserve

CREATING A COMPETITIVE ADVANTAGE

Revenue Driven

1. Increase viewer engagement

- Longer time spent viewing content on platforms
- Increase advertising impressions

2. Change viewer behavior

- Change from passive viewer to engaged viewer
- Engage in a fair exchange of data for a better experience
- Be willing to accept recommendation. Build trust on content

CREATING A COMPETITIVE ADVANTAGE

Revenue Driven Cont....

3. Reactivate dormant viewers

- Have the ability to make tailored recommendations to viewers who signed up to watch a specific piece of content
- Drive platform adoption and position provider as a source of entertainment across multiple genres

Customer Experience Driven

1. The ability to identify viewers by persona along with associated content tastes, preferences, and viewing habits
2. The ability for business users to glean key insights from data through intuitive visuals and other mechanisms that enable them to make business decisions



HAIL
HYBRID.

BOHEMIAN RHAPSODY

135m | Drama Music

Singer Freddie Mercury, guitarist Brian May, drummer Roger Taylor and bass guitarist John Deacon take the music world by storm when

SEE DETAILS

Matrix Factorization Filter 1.70258359

Bohemian Rhapsody | Teaser Trailer [HD] | 20th Century FOX

BOHEMIAN RHAPSODY

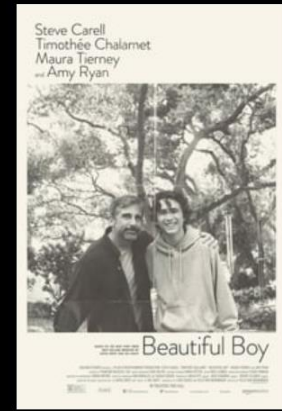
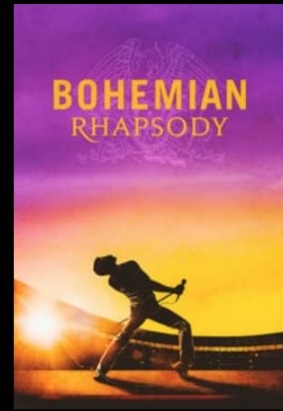


OFFICIAL TRAILER



Watch later Share

RECOMMENDED FOR YOU



How Does It All Work?

Recommendation Engine Components

Mixture of Recommenders (Unlimited)

- Content-Based
- Collaborative-Based
- Hybrid

User Profile

- User Feedback
- User Interaction

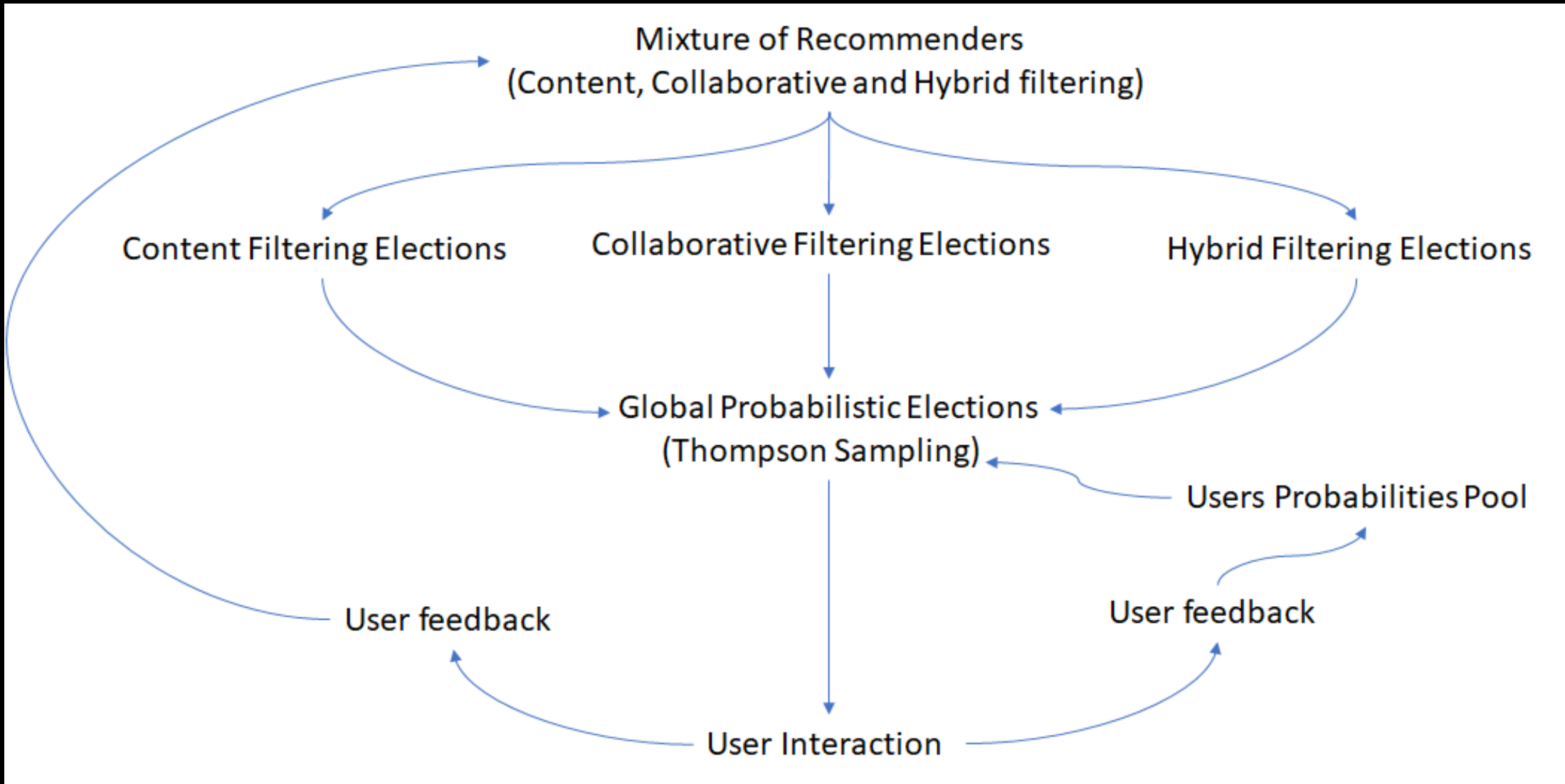
Global Probabilistic Electors (Thompson Sampling)

THE MATH BEHIND IT ALL

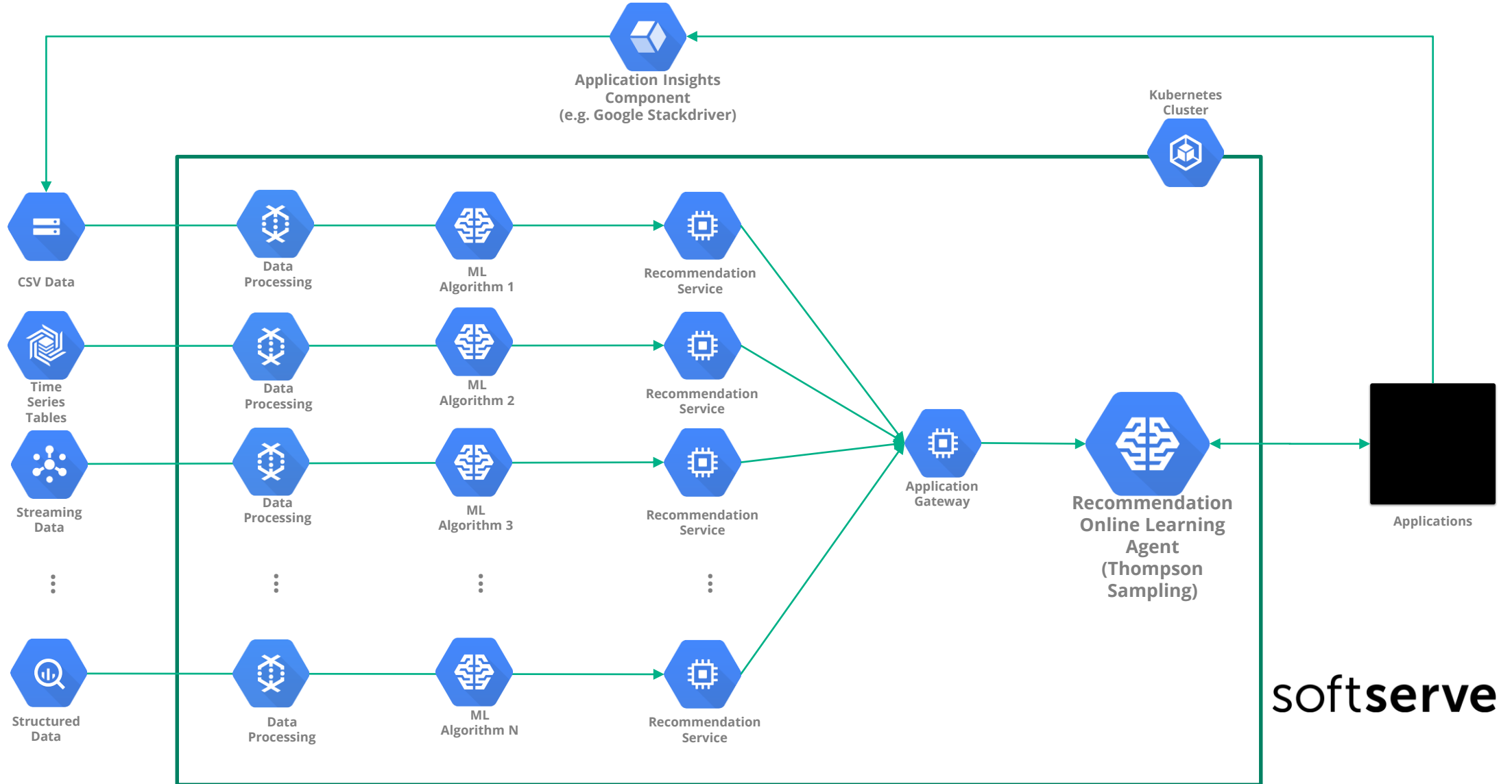
Thompson Sampling

- In artificial intelligence, Thompson sampling is a heuristic for choosing actions that addresses the exploration-exploitation dilemma in the multi one-armed bandit problem
- It consists in choosing the action that maximizes the expected reward with respect to a randomly drawn belief
- It has been proved that Thompson sampling is instantaneously self-correcting.

$$\int_{\theta} \max_{a'} r(a', x, \theta) P(\theta) \mathcal{D}_t d\theta,$$



DATA SCIENCE ALGORITHMS IN ACTION



WHY HAIL HYBRID?

Hybrid Approach

- The engine can evaluate an unlimited number of data sources used in the recommenders – machine learning algorithms that compare data sets, user actions, social activity, retail data, etc.
- All of these various recommenders are processed and sent to the Recommendation Learning Agent to use probability sampling algorithms in order to find the most likely matches
- As this process is repeated over time, the data and machine "knowledge" becomes even more accurate and rewarding

MAKING IT PERSONAL

User Insights

- User feedback such as ratings and likes
- Direct user activity with the application/recommendation engine
- Consumption habits – duration and time of day
- Social data
- Retail data
- Geographic/demographic data

MAKING IT OPEN & SCALABLE

Additional Highlights

- Taking a containerized approach allowed extreme burst capacity enabling us to chunk through huge amounts of data in micro-seconds and even nano-seconds
- The architecture is built on an open source codebase and open and with open source technologies. The recommendations engine framework does not need to be purchased or licensed. It is also cloud agnostic
- The inputs are unlimited and customizable. You can perfect the engine to your customer base or target viewership based on data inputs and algorithm calculations

KEY TAKEAWAYS

- An incredible amount of data is available: Put it to use
- Transparency and Trust: People want to know why recommendations were made and what data was used
- You'll need a test group: Be nice to your friends and co-workers

**FOR
THE
FUTURE**

softserve