The Big Book of Media & Entertainment Use Cases
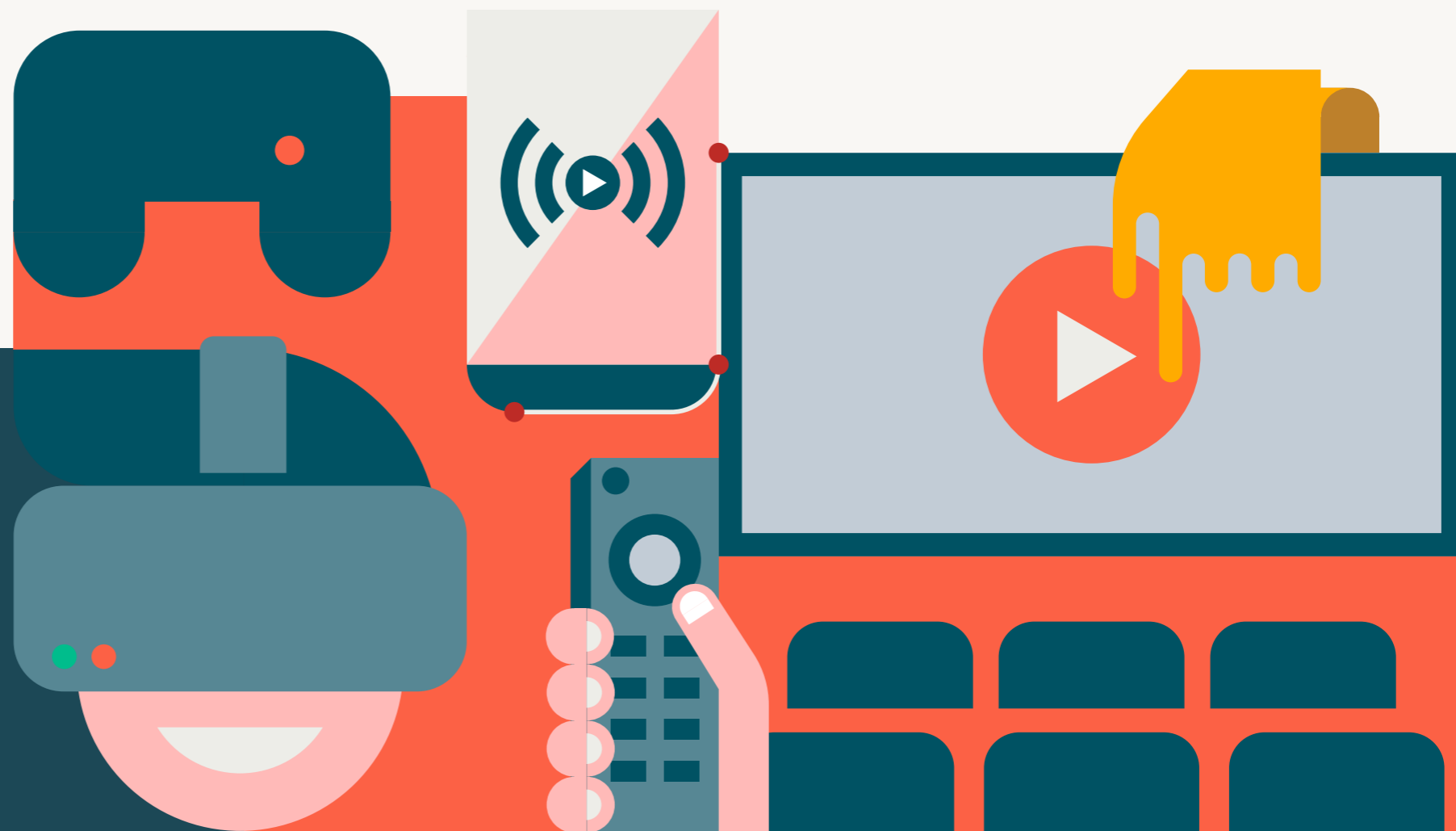
Putting data at the center of consumer and advertiser experiences

databricks

Contents

databricks

Lakehouse for Media & Entertainment

Chief data officers and data leaders in the media and entertainment industry have more choices than ever before when it comes to building next-generation analytics platforms. Unique in use cases compared with other industries, media and entertainment is inherently a creative business where the ability to handle unstructured data such as image, video and audio is mission-critical to being successful with analytics workloads. Combined with rising consumer expectations for real-time, frictionless and personalized experiences, many organizations struggle with cost, performance and reliability in scaling data pipelines that enable AI and machine learning to be at the heart of their consumer and employee experience.

In this use case guide, we show how the Databricks Lakehouse for Media & Entertainment is helping leading media organizations take **all their data in a single architecture, make it ready for SQL and AI/ML, and do so quickly within their own cloud infrastructure environment based on open source and open standards**. From handling massive scale streaming workloads to having the ability to handle unstructured data, the Databricks Lakehouse for Media & Entertainment is sure to be the standard for how media and entertainment organizations operationalize their analytics efforts. This use case guide also highlights common use cases across the industry, from personalization and consumer experience to driving better outcomes for advertising-based companies. Resources in the form of Solution Accelerators and reference architectures are also included to help you as you embark on your own journey to drive better outcomes with big data and AI.

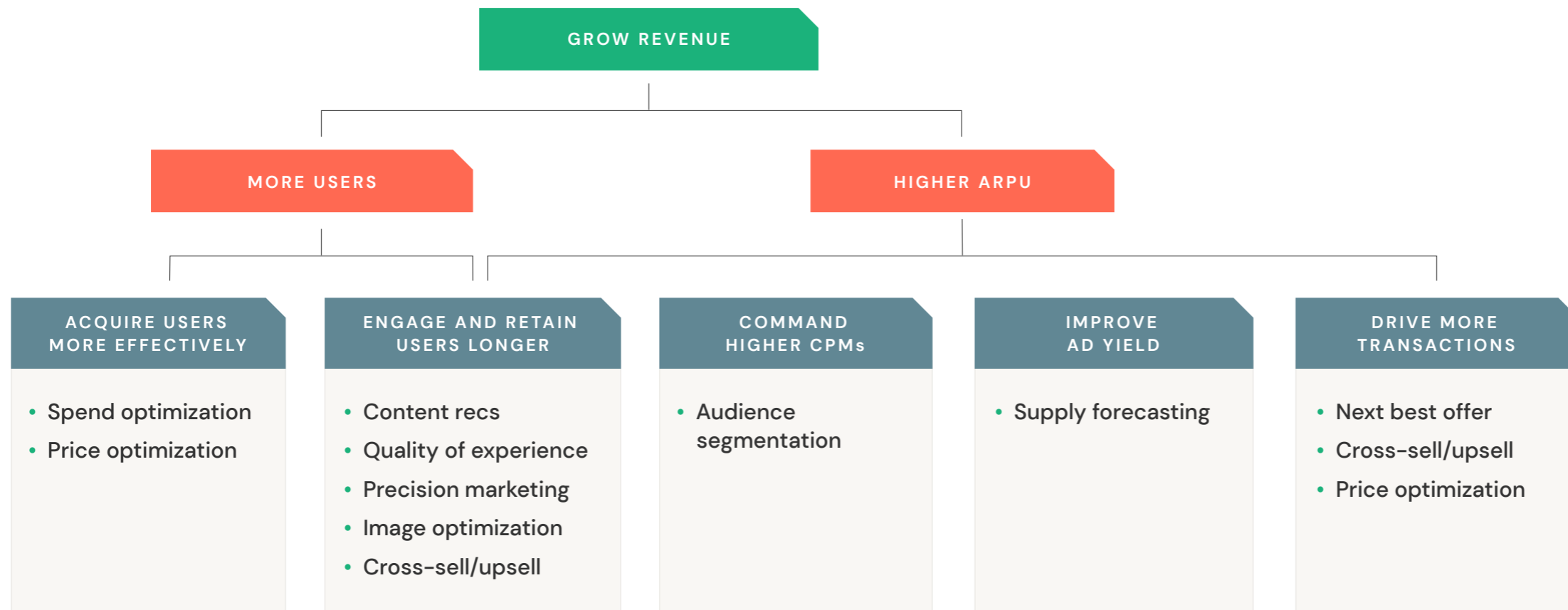## Customers That Innovate With Databricks Lakehouse for Media & Entertainment

Some of the top media and entertainment companies in the world turn to Databricks Lakehouse to bring AI-driven innovation and excellence to their customers.

A&E   CBC Radio-Canada   CONDÉ NAST   COMCAST   Dolby

NBCUniversal   Nielsen   SEGA   SHOWTIME   Paramount   zynga

databricks

Modern Cloud Data Infrastructure

Overview

Companies across the media industry continue to adapt to new business models and experiences that align with changing consumer preferences. In navigating the change, much focus is now on initiatives that drive revenue through the growth and monetization of a highly engaged user base.

GROW REVENUE

MORE USERS

HIGHER ARPU

ACQUIRE USERS MORE EFFECTIVELY

- Spend optimization
- Price optimization

ENGAGE AND RETAIN USERS LONGER

- Content recs
- Quality of experience
- Precision marketing
- Image optimization
- Cross-sell/upsell

COMMAND HIGHER CPMs

- Audience segmentation

IMPROVE AD YIELD

- Supply forecasting

DRIVE MORE TRANSACTIONS

- Next best offer
- Cross-sell/upsell
- Price optimization

databricks

Common challenges

As companies execute these initiatives, they often find that the state of their data hinders them from moving with the speed and agility that the market demands. A few common roadblocks include:

Moving beyond aggregation (to advanced analytics)

Prior to using analytical techniques such as media mix modeling for spend optimization, or survival analysis for churn mitigation, a big lift is often needed to acquire and harmonize a large number of data sets. In some cases, this work requires a capital investment and cross-team coordination.

Creating unified audience profiles

Audience data has traditionally been captured, stored and managed directly in disparate systems (e.g., DMP, ESP, data lake, data warehouse) based on size and granularity, intended use case(s), and data types. This siloed approach increases the complexity of managing customer data as an asset that can be used to support a variety of use cases (e.g., content recommendations, next best offer).

Delivering a flawless user experience

A by-product of media consumers having more choice than ever before is that delivering a flawless user experience is now merely table stakes. At the same time, doing so requires being able to identify the quality of service issues in near real-time, a capability that is not directly supported by the existing tech stack at many companies.

Unlocking the value of media assets

In many media companies, the largest source of data is not clickstream or ad server logs, it's the back catalog of content. Whether it's to improve content recommendations, command higher CPMs through custom audience segments or facilitate the repackaging of content for pop culture events, the value of combining media assets with structured data, and then feeding the result into ML/AI algorithms, cannot be overstated. With media assets managed in isolation, however, many companies struggle to realize such benefits.

Deploying machine learning models to production

Whether it's to personalize the user experience or to equip your customer support team with the next best offer for a subscriber at risk, a machine learning model doesn't realize its full value until it's deployed into production. However, deployment is often slowed down due to a required handoff between teams, and once deployed, the model is often at risk of becoming unstable because it's built using a data source that does not enforce data quality.
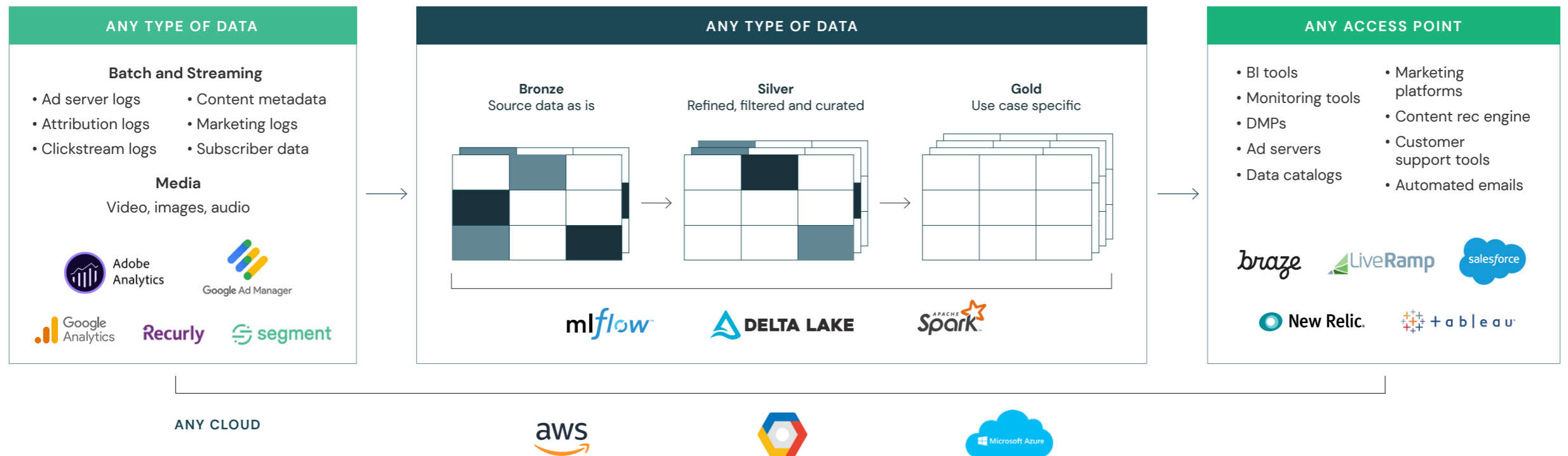
databricks

Solution overview

While it's possible to address these types of roadblocks in a business-as-usual manner, doing so won't address the actual factors that are hindering speed and agility long term. What's required is a modern data architecture that reduces complexity, increases flexibility and fosters reusability. Enter the lakehouse. The lakehou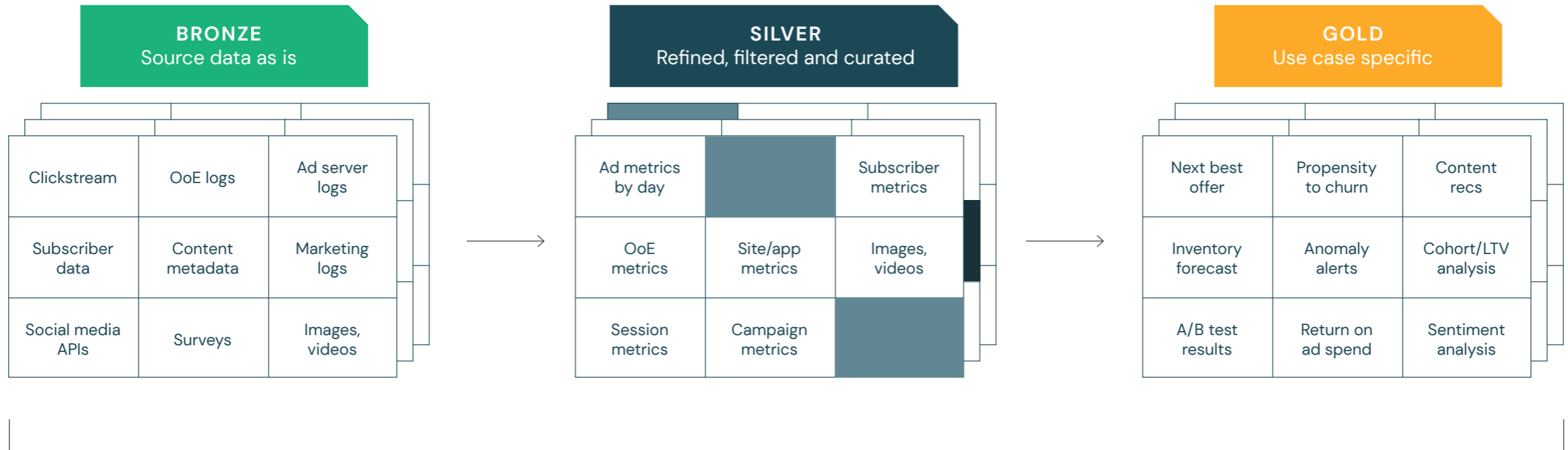se — a new paradigm that combines the best elements of data warehouses and data lakes — directly addresses these factors by enabling you to unify all your data and then support all your use cases in a consistent way. With your data stored in an open format, on a low-cost object store and in your own cloud, you also get a cost-effective solution that guarantees that you maintain full ownership and control over your data.

Reference architecture

SYSTEM VIEW

DATA VIEW

BRONZE
Source data as is

Clickstream	OoE logs	Ad server logs
Subscriber data	Content metadata	Marketing logs
Social media APIs	Surveys	Images, videos

SILVER
Refined, filtered and curated

Ad metrics by day		Subscriber metrics
OoE metrics	Site/app metrics	Images, videos
Session metrics	Campaign metrics	

GOLD
Use case specific

| Next best offer	Propensity to churn	Content recs
Inventory forecast	Anomaly alerts	Cohort/LTV analysis
A/B test results	Return on ad spend	Sentiment analysis

mlflow

DELTA LAKE

APACHE Spark

databricks

How it works

The reason that the lakehouse is able to bring the simplicity, flexibility and reusability required for you to move at the speed and agility that the market demands is that it is fundamentally a reimagining of the modern data architecture.

▶ Any type of data can be stored because, like a data lake, the lakehouse is built using the low-cost object storage supported by cloud providers. Leveraging this capability helps break down the data silos that hinder efforts to aggregate data for advanced analytics (e.g., media spend optimization, next best offer), create and monetize custom audience segments using first-party data, and unlock the value of media assets (e.g., video, images, audio, text).

▶ Streaming use cases such as monitoring video quality of service are simpler to support because the lakehouse uses Apache Spark™ as the data processing engine and Delta Lake as a storage layer. With Spark, you can toggle between batch and streaming workloads with just a line of code. With Delta Lake, native support for ACID transactions means that you can deploy streaming workloads without the overhead of common issues involving reliability (e.g., jobs failing midway) and performance (e.g., "too many files").

▶ Machine learning models, such as a content recommendation engine, can be deployed to and supported in production more easily because the lakehouse uses Delta Lake and MLflow. With Delta Lake, you can ensure your machine learning models remain stable by declaring your expectations for data quality upfront. In the context of content recommendations, this can be used to prevent upstream data collection issues from compromising the user experience. With MLflow, you can train your model in any language and deploy it anywhere, without the extra overhead caused by integration work. This minimizes the need for a cumbersome "handoff" between a data scientist and a software engineer.
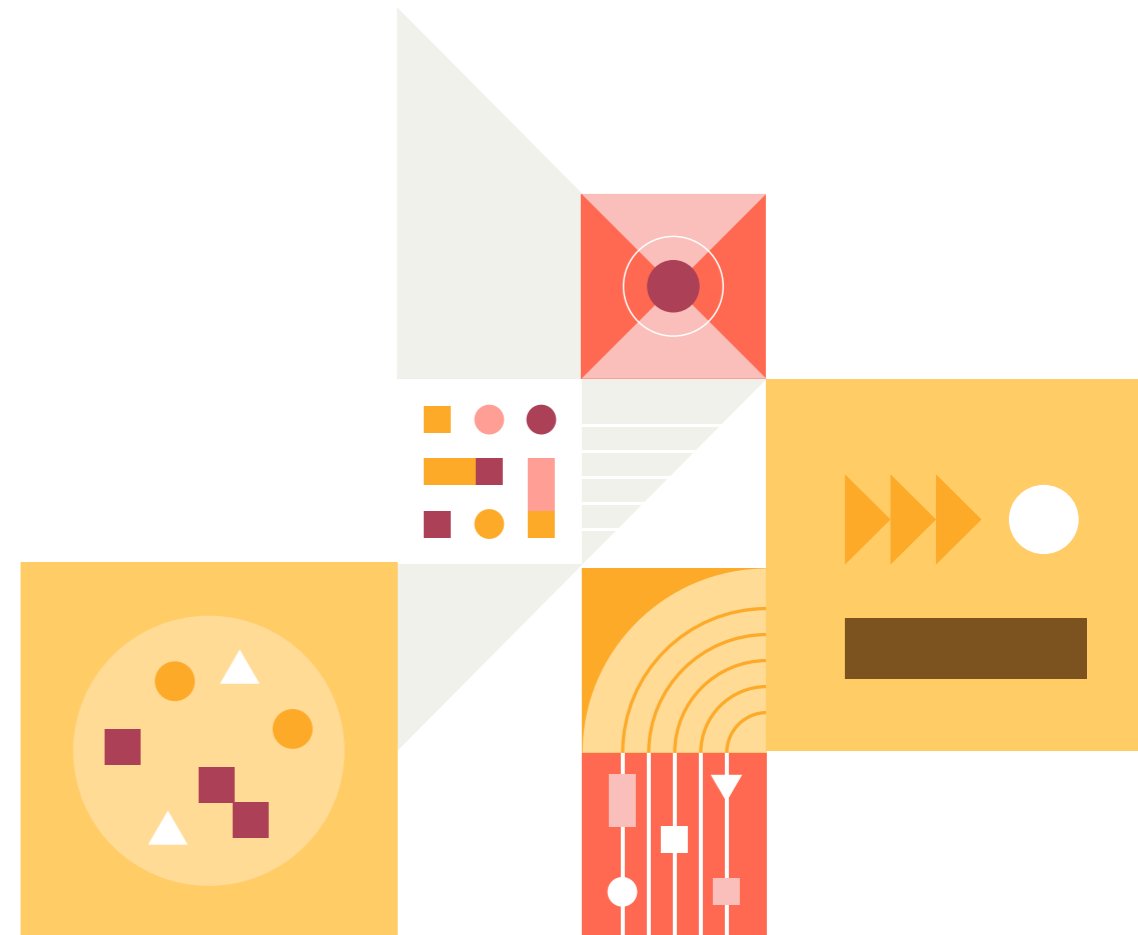
databricks

Value with Databricks

By using Databricks Lakehouse for Media & Entertainment to build and support your lakehouse, you can equip your business with even more speed, agility and cost savings.

Collaborative workspace
Databricks provides data analysts, scientists and engineers with a common workspace offering purpose-built tools that promote collaboration over handoffs. For example, a data scientist can get started on an algorithm for marketing spend optimization while a data engineer acquires and harmonizes new data sources and a marketing data analyst provides the requisite domain knowledge.

Integrated platform
Databricks is fully integrated with all of the components of the lakehouse architecture. This integration helps free up resources so you can focus on solving business problems instead of hosting software.

Reduced infrastructure costs
Databricks is deployed on your existing cloud infrastructure and provides you with the tooling needed to manage your compute resources (e.g., auto-scaling, auto-terminate) effectively. For example, in the context of quality of experience, you can easily scale up your computing resources to handle peaks in traffic (e.g., evening binge-viewing) and then scale back down once appropriate.

Getting started

The flexibility of the lakehouse means that you can start with the use case that will be most impactful for your business. As you implement the pattern, you will find that you're able to tackle use cases quicker and more easily than before. To get you started, this guidebook contains the use cases we most commonly see across the media and entertainment industry. A few of these use cases are listed below.

Campaign optimization
Leverage multi-touch attribution to continuously optimize spend within and across campaigns.

Content recommendations
Harness the value of unstructured data (e.g., video, images, audio, text) to drive engagement through content recommendations.

Customer 360
Construct and leverage audience profiles to mitigate churn and boost LTV through personalization, messaging and product development.

Quality of experience
Identify and resolve quality of service issues in near real-time.

Yield, pricing and inventory
Forecast advertising inventory at a granular level to effectively manage supply and demand, minimize waste and command premium CPMs.

databricks

Reference architecture

Below is a common reference architecture for where Databricks Lakehouse for Media & Entertainment fits within a media and entertainment data estate. Databricks commonly handles both batch and streaming workloads as part of the ingestion process, and the data is then refined by a Bronze, Silver and

Gold schema that makes data ready for downstream BI or AI/ML use cases. Common egress sources could be BI solutions for data visualization or routing solutions that help deliver experiences to ad tech, mar tech or other consumer–facing platforms.
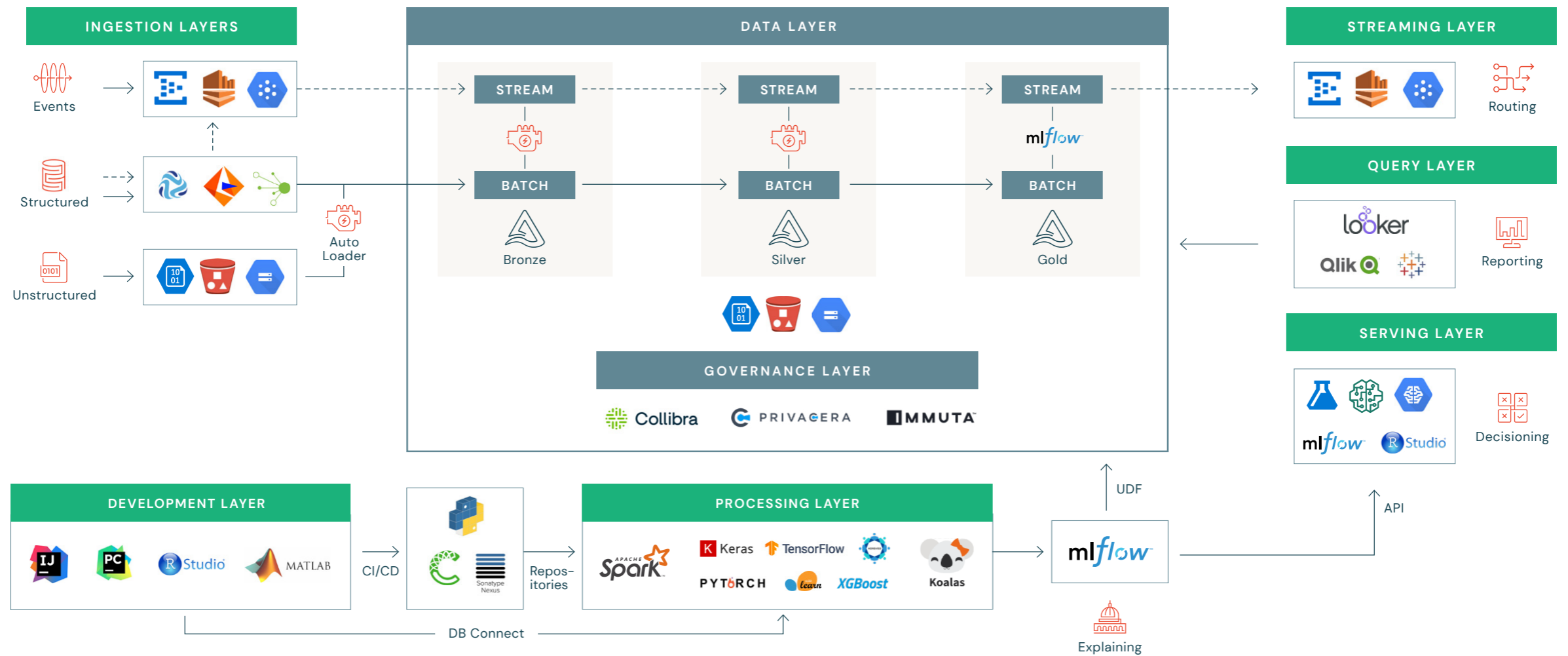
USE CASE: **Quality of Experience**

Overview

A key factor in improving consumer retention and engagement, quality of experience is a measure of a content provider's service quality from the consumer's point of view. Poor buffering, network service and overall performance are all factors that lead to end users having a less than satisfactory experience on a direct-to-consumer media platform.

Relevant for

Streaming video and audio services, subscription video on demand, advertising video on demand, digital natives and gaming.

Challenges

User experience
Videos failing to load or constant buffering issues.

Audience growth
Massive amounts of session data to mine for audience insights.

Targeting advertising
Ad sales/conversions failing due to limited audience engagement.

Value with Databricks

Databricks Lakehouse for Media & Entertainment has enabled this use case across a variety of video streaming services. Industry benchmarks include:

• **Reduced video start delay by 33%**
Predict trends and streaming issues to provide a superior viewing experience.

• **Increased viewer retention by 3.5x–7x**
Leverage viewership data to understand best strategies to drive engagement.

• **Increased ad conversions**
Target customers with personalized ads based on Comscore ratings and viewer behavior.

Solution overview

Databricks allows for both streaming and batch data sets to be analyzed to ensure a performant streaming content experience. Delta Lake enables the ability to handle real-time and near real-time use cases as well as the workflow for any experience degradations, which is critical to scaling this use case. Typical use case data sources include:

• **Asset:** video title, ad creative name, rendition, content length, content ID, CMS
• **Device:** user agent, device type, device model, operating system, device/user ID
• **Geo:** IP address, location (e.g., lat/long, city, state, country)
• **Delivery:** ISP, CDN, connection type (e.g., Wi-Fi), video player
• **Metrics:** video starts, time to the first frame, video start failure, playback failure, rebuffering ratio, bit rate, session length

databricks

How to get started

Read our how-to blog, which includes notebooks you can leverage to build your own solution.

> **How to build a Quality of Service (QoS) analytics solution for streaming video**

As consumers continue to choose streaming services over traditional television, getting ahead of video quality issues (encompassing buffering, latency, pixelation, jitter, packet loss and the blank screen) is a key part of decreasing subscriber churn and increasing video engagement.

> **QoS notebook**

Get hands-on with an end-to-end project using Delta Lake and a Delta architecture pattern in these provided notebooks. They showcase how to ingest data with real-time data enrichment and anonymization, set up real-time notifications using rules or ML-based scoring, update the web application in real time with real-time aggregations, and run your network operations center with a dashboard that can be shared quickly.

Case Study

Paramount

CUSTOMER STORY >

With Databricks Lakehouse for Media & Entertainment, Paramount constantly monitors the quality of video feeds and reallocates platform resources in real time to ensure high quality of service.

databricks
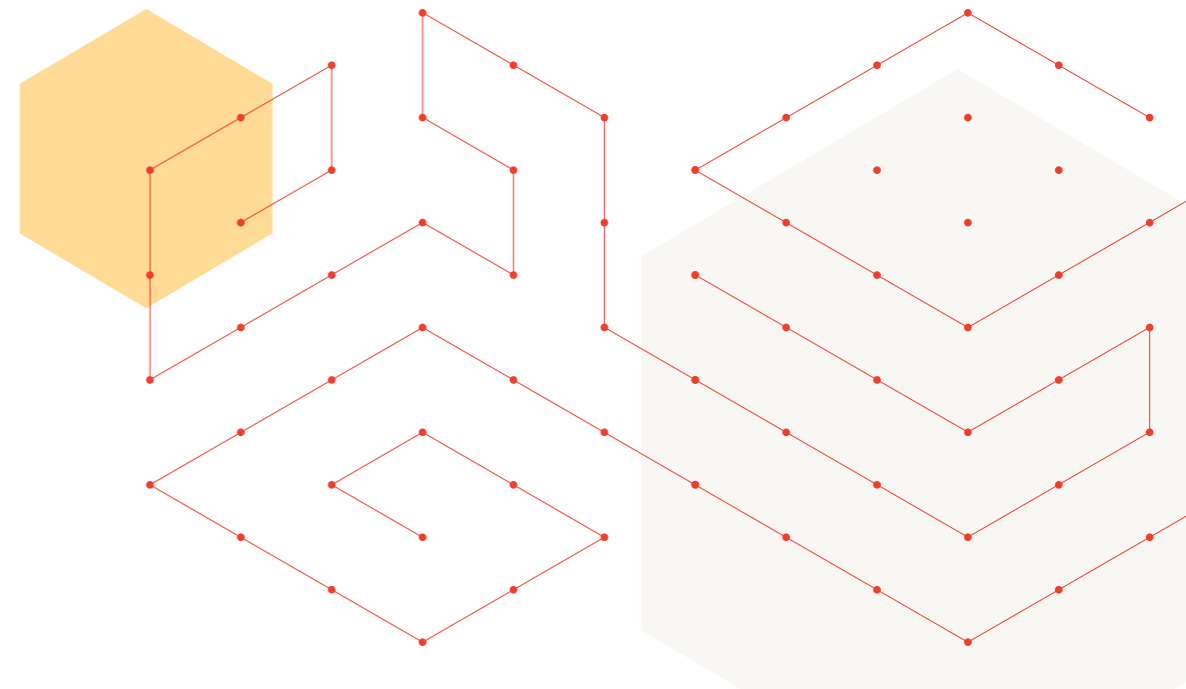
USE CASE: # Churn/Survivorship

Overview

The media and entertainment industry has pivoted heavily from wholesale to retail, with direct-to-consumer products now driving revenue growth across many media companies. Churn management has now become one of the largest issues facing the D2C market as customers leave services based on seasonality, content, price or simply not having a relevant, tailored experience. Most organizations understand backward-looking metrics (e.g., Who churned? How much revenue will I lose?), but media companies must now look at forward-looking metrics including churn prediction, revenue loss estimation, churn context and how to intervene to prevent subscriber loss. Using Kaplan-Meier curves and Cox Proportional Hazard models, this use case identifies customer behaviors to predict increased risk of subscription cancellation.

Relevant for

All direct-to-consumer media companies including gaming/gambling, publishing, streaming video and audio services, and others selling subscription services.

Challenges

Cost of acquisition

Customer retention is up to 80% cheaper than customer acquisition, so predicting and preventing churn has high economic value.

Revenue linearity

Measuring direct-to-consumer services on active user counts and unanticipated churn has short-term financial implications and long-term share value consequences.

Customer expectations

Personalization is now an expectation for consumers, with churn driven higher by services that aren't focused on 1:1 experiences.

Scale of data

Subscriber data has many features and elements where sampling is ineffective unless it's looking at the entirety of the data set.

Value with Databricks

Databricks Lakehouse for Media & Entertainment has enabled this use case across a variety of publishers, gaming companies and streaming video platforms. Sample value metrics include:

- **6x faster pipelines**
  Data pipelines that took over 24 hours are now run in less than 4 hours, enabling teams to make decisions faster.

- **Removing infrastructure complexity**
  A fully managed platform in the cloud with automated cluster management allows the data science team to focus on machine learning rather than configuring hardware, provisioning clusters, debugging, etc.

- **Innovating the subscriber experience**
  Improved data science collaboration and productivity have reduced time-to-market for new models and features. Teams can experiment faster, which leads to a better, more personalized experience for subscribers.

databricks

Solution overview

The key to successfully completing this work is the establishment of transparent, maintainable data processing pipelines executed on an elastically scalable (and therefore cost-efficient) infrastructure, a key driver behind the Delta Lake pattern. While most organizations may not be overly cost-conscious in their initial approach, it's important to remember the point made above that churn is a chronic condition to be managed. As such, this is an analysis that should be periodically revisited to ensure acquisition and retention practices are aligned. Typical use case data sources vary by segment but will include:

- **Telco:** plan type, auto-pay, family plan, cohort (e.g., heavy video, casual gamer, email and web)
- **Media:** subscription plan, "fan of," cohort (e.g., comedy fan)
- **Gaming:** cohort (e.g., evening player), last level completed
- **General:** acquisition source
- **Metrics:** subscription length, propensity to churn, video views, gameplays, high score, days since last use, in-app purchase transactions/revenue

How to get started

> **Analyzing Customer Attrition in Subscription Models**

In this blog post, you will learn how to use the Delta Lake pattern to apply churn analytics use cases to your batch and streaming data sets.

SURVIVAL ANALYSIS NOTEBOOKS

There are four notebooks available to support this use case, from data prep through moving the use case into production. These notebooks include:

- Survival Analysis 01: Data Prep
- Survival Analysis 02: Exploratory Analysis
- Survival Analysis 03: Modeling Hazards
- Survival Analysis 04: Operationalization

Case Study

SHOWTIME.   CUSTOMER STORY >

With the Databricks Lakehouse Platform, Showtime now has an actionable view into the consumer journey, with the goal of increasing engagement while lowering churn.

USE CASE: **Next Best Offer**

Overview

One of the largest issues for media companies with subscription models is saving customers who say they want to cancel a subscription. This is often done with a new offer that will fully or partially recoup any expected revenue loss. Some companies overdiscount new offers, which leads to a decrease in product profitability and overall customer lifetime value. Other companies don't use contextual user-level data to understand what offers will lead to the highest conversion rate. Generating the right offer at the right time is critical for agents trying to upsell or save a customer in a call center environment, as well as presenting the right offer on self-assisted channels or outbound marketing. This use case examines how to build a real-time engine to present the right offer to save, win back or convert with optimized offer management.

Relevant for

All direct-to-consumer media companies including gaming/gambling, publishing, streaming video and audio services, and others selling subscription services.

Challenges

Segmentation
Dynamically and continually grouping consumers into micro-segments is necessary to determine who will behave similarly and have similar responses to actions.

Real-time machine learning
Ability to test, track and optimize ML lifecycle based on how marketing actions are affecting consumer micro-segments.

Flexibility to handle behavior modeling
Using ML to predict how much each micro-segment of consumers will respond to each available marketing action.

Value with Databricks

Databricks Lakehouse for Media & Entertainment has enabled this use case across a variety of publishers, streaming video platforms and communication service providers. Sample value metrics include:

- **50% reduction** in big data processing time

- **35% reduction** in big data processing costs from legacy infrastructure solution

databricks

Solution overview

Using Delta Lake as the foundation for personalization–based use cases, transactions can be joined with customer demographics, payments, ratings and other data sources to have a holistic view of the customer. Time travel is used to ensure full reproducibility of each propensity model and MLflow to rapidly prototype, evaluate and release those models. Finally, the ability to combine multiple models across many products and services into a common decisioning engine helps organizations reduce churn risk while maintaining superior customer experience and acquiring new customers and retaining existing customers with a next best offer. Typical use case **data sources** include:

- Subscription plan type, subscription length, cost of acquisition, lifetime value, propensity to churn

- Video views, gameplays, billing history, genre affinity, date of last use, engagement score, next offer recommendations

How to get started

Solution Accelerators for Media & Entertainment

Based on best practices from our work with the leading M&E brands, we've developed Solution Accelerators for common analytics and machine learning use cases to save weeks or months of development time for your data engineers and data scientists. Visit this site for regular updates on new use cases including next best action/offer.

Case Study

dish

CUSTOMER STORY >

Dish is among many of the digital video services Databricks works with that has built a next best offer use case to help save customers at risk and maintain customer lifetime value. Learn more about Dish and others at our media overview page.

databricks

>

USE CASE: **Content Recommendations**

Overview

Personalization is now table stakes for any direct-to-consumer company that is looking to drive consumption, brand affinities and the ability to increase customer wallet share. One aspect of personalization is ensuring that relevant content recommendations are surfaced to end users based on their behavior, preferences, similarity to other consumers and other attributes. One of the challenges of content recommendation systems is the ability to drive real-time or near real-time use cases with limited consumer engagement or behavior that does not follow other users. This use case examines the various approaches that can be taken in building out and scaling a content recommendation system.

Relevant for

Direct-to-consumer subscription platforms including video, audio, publishing, broadcasting and gaming as well as advertising-supported platforms looking to drive a personalized experience.

Challenges

The scale of data
The most effective content recommendation engines are taking all clickstream data to understand dwell time and other behavioral metrics.

Dynamic user data
Data tables will need to constantly be appended based on any changes to preferences or other elements of user data.

Real-time analytics
User cold-start is often a huge challenge for all direct-to-consumer services.

Value with Databricks

With Databricks Lakehouse for Media & Entertainment as the foundation for their data analytics and machine learning efforts, media companies have realized:

- **Improved customer engagement**
  With an improved data pipeline, media companies can make better, faster and more accurate content recommendations, improving the user experience.

- **Unified approach**
  Data engineering and data science teams can now solve problems together and collaborate to build new content products and experiences.

- **Built for scale**
  Data sets can no longer outgrow capacity to process and glean insights.

- **More models in production**
  With MLflow, data science teams can innovate their products faster. A sample customer in the publishing segment has over 2,000 models in production on Databricks.
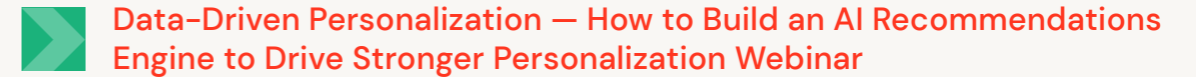
databricks

Solution overview

Databricks Lakehouse for Media & Entertainment enables organizations to combine various sources of data in a timely and efficient manner, including batch and streaming data sources that are needed to get a full view of a consumer's behavior which can be used for downstream recommendations. MLflow is also a critical component of this use case. By applying data science techniques on top of consumer behavior indexed against content metadata, it enables the recommendation process. Typical use case data sources include:

- **Content:** series, episode, genre, premiere/release date, director, producer, cast, geo rights

- **User/device:** user ID, geo, IP address

- **Slate:** slate position

- **Metrics:** video starts, minutes viewed, percent viewed

How to get started

Personalizing the Customer Experience With Recommendations
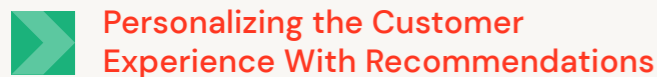
In this blog post, learn more about some of the factors you need to consider when building recommender systems.

Data-Driven Personalization — How to Build an AI Recommendations Engine to Drive Stronger Personalization Webinar

View this on-demand workshop to see best practices and techniques for building recommendation systems.

Recommender Notebooks
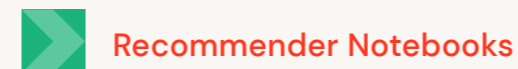
Leverage these Solution Accelerator notebooks to explore what it takes to build both collaborative filtering and content-based recommenders.

Case Study

CONDÉ NAST   **CUSTOMER STORY >**

Databricks Lakehouse for Media & Entertainment provides Condé Nast with a fully managed cloud platform that simplifies operations, delivers superior performance and enables data science innovation. Delivering a personalized experience across their brands is a major focus of how they use Databricks to help create 1:1 relationships with millions of consumers.

databricks

>

USE CASE: Customer Lifetime Value

Overview

Understanding and identifying who your most valuable customers are will help guide better marketing investment and product development choices. As media companies shift more to direct-to-consumer models, aligning customer acquisition cost (CAC) and retention rates with the total revenue or customer lifetime value is critical in ensuring that services are profitable. By calculating the amount of revenue received from a given customer over the lifetime of our relationship with them, we might better tailor our investments to maximize the value of our relationship for both parties. This use case focuses on retention and spending components to then combine into an overall CLV model ideal for measuring estimated lifetime and estimated customer spend for a transaction-based business like TVOD or AVOD as well as subscription-based businesses.

Relevant for

All direct-to-consumer media companies including gaming/gambling, publishing, streaming video and audio services, and others selling subscription services. Variations of CLV can also be used for advertising and content sales businesses.

Challenges

Volume of data

A large volume of transaction-based data must be processed in order to generate per-customer metrics required by many CLV models.

Model iteration

Curves must be derived from this data to fit into expected patterns of value distribution. This process is regulated by a parameter that cannot be predetermined and instead must be evaluated iteratively across a large range of potential values.

Aligning to marketing lifecycle

To maximize the effectiveness of CLV models, there must be integration with marketing and customer engagement activities to ensure the model generates meaningful impact.

Value with Databricks

With Databricks Lakehouse for Media & Entertainment as the foundation for their determining CLV, media companies have realized:

- **Scale**
  Databricks can handle massive amounts of data — whether structured or unstructured — and can scale to calculate customer metrics for each individual.

- **Performance**
  Databricks excels at bringing critically necessary functions to this use case, such as optimized hyperparameter tuning across a large search space.

- **Experimentation**
  The ability to iterate and deploy an optimal model quickly is a key value that Databricks delivers relative to this use case.

databricks

Solution overview

Leveraging MLflow, a machine learning model management and deployment platform, we can easily map our model to standardized application program interfaces. The end result of this is that we can quickly turn our trained models into functions and applications enabling periodic, real-time and interactive customer scoring of life expectancy metrics. Typical use case data sources include:

- **Acquisition source, cohort**

- **Metrics:** ad revenue, subscription revenue, transaction revenue, cost of acquisition, subscription length

How to get started

Estimating Lifetime Duration

In this how-to blog post, which includes links to notebooks that you can use to build your own CLV models, the Buy Til You Die approach is leveraged to show how to derive an estimated customer lifetime.

Estimating Future Spend

This blog post shows how users can build a model that calculates the monetary value associated with future purchase events and combine it with lifetime probabilities to determine estimated customer lifetime value.

Maximizing Customer Value

This practitioner-focused webinar goes deep into both of the notebooks listed above and walks through common user questions in establishing CLV models.

Case Study

SEGA®   **CUSTOMER STORY >**

EGA leverages the Databricks Lakehouse for Media & Entertainment to deliver CLV use cases, unifying massive amounts of data to derive insights that deliver personalized experiences at scale to 1 million gamers daily.

databricks

>

USE CASE: **Omnichannel Personalization/Customer 360**

Overview

Omnichannel personalization is the process of tailoring personalized experiences for users across channels based on the data collected from the users' behavior on other digital channels. It is channel agnostic and encompasses email, SMS, authenticated web, social and other channels to create a 1:1, contextual experience. Leveraging a customer 360 architecture that collects all user clickstream and behavioral data, marketers are able to create customized experiences that work across business imperatives like churn reduction, user satisfaction and user engagement. This use case focuses on how customer 360 is the enabling architecture that allows marketers to drive personalized experiences across their marketing execution stack, with Databricks Lakehouse for Media & Entertainment as the staging area for customer data and applied machine learning to deliver personalization.

Relevant for

All marketing organizations and data teams responsible for consumer data. Of particular interest to direct-to-consumer services looking to drive customer engagement.

Challenges

Identity and comprehensive data
Data silos prevent the ability to identify and stitch together all customer touchpoints in one place to drive accuracy and precision in messaging.

Processing speed
Real-time processing is required for this use case, as timely messages ensure that the underlying personalization is effective.

Automation
This is an "always-on" use case where automation is essential for scalability and responsiveness based on frequent model updates.

Value with Databricks

With Databricks Lakehouse for Media & Entertainment as the foundation for their customer 360 architecture to deliver omnichannel personalization, sample value metrics from a media agency include:

- **Revenue uplift**
  45%–50% increase in campaign revenue for a sample consumer goods client focused on multichannel personalization.

- **Infrastructure cost savings**
  22% lower operational cost of running their data pipelines year over year.

- **Increased data engineering/data science collaboration**
  30% productivity improvement across data teams. The ability to iterate and deploy an optimal model quickly is a key value that Databricks delivers relative to this use case.

databricks

Solution overview

Databricks Lakehouse for Media & Entertainment enables organizations to combine various sources of data in a timely and efficient manner — from transactions, demographics and preference information across products to clickstream, digital journey and marketing analytics data to bring a 360 view of customer interactions to enable omnichannel personalization. By identifying changes in user behavior or engagement, media companies are able to detect early signals for churn by proactively engaging and recommending content and services that will keep consumers sticky across subscription or ad-based platforms. Typical use case data sources include:

- **Identity:** household ID, person ID, device ID, email, IP address, name, gender, income, presence of children, location
- **Engagement:** affinities (show, games, magazines, genres), cohorts/traits, date of last use, QoE events (e.g., app crash, video playback error)
- **Messaging:** message preference (e.g., SMS, email), message history
- **User lifecycle:** subscription status (e.g., free trial, paid subscriber), payment history, cost of acquisition, lifetime value, propensity to churn

Case Study

PUBLICIS GROUPE

CUSTOMER STORY >

With Databricks Lakehouse for Media & Entertainment, Publicis Epsilon has enabled its retail clients to convert prospects into customers quickly and retain them for longer periods of time, resulting in increased campaign revenue by as much as 50%.

How to get started

To see how Publicis Epsilon architected an omnichannel personalization engine with Databricks Lakehouse for Media & Entertainment, check out the following customer presentation:

Deliver Dynamic Customer Journey Orchestration at Scale

As customer acquisition costs continue to rise steadily, organizations are looking into ways to optimize their end-to-end customer experience in order to convert prospects into customers quickly and to retain them for a longer period of time. In today's omnichannel environment where nonlinear events and micro-moments drive customer engagement with brands, the traditional one-size-fits-all customer journey will not be able to deliver true value to the customer or to the organization. Trained on one of the largest customer data sets available in the United States, Publicis Epsilon leverages Databricks Lakehouse for Media & Entertainment to stitch together and analyze profile-based, behavioral, transactional, financial and operational data to deliver customer journey orchestration at scale.

Best Practices for CDP Design and Implementation Using Databricks and AWS

Many marketing and technology organizations struggle to make the right decision when approaching how to leverage a consumer data platform (CDP) to help orchestrate their segmentation and marketing activation strategy. In this joint blog with AWS, we explore best practices for building or augmenting consumer data platforms (CDPs) with Databricks Lakehouse for Media & Entertainment and AWS, including what recommended approaches are best suited for your organization based on size and technical competency.

databricks

USE CASE: # GDPR/CCPA Compliance

Overview

With more digital data being captured every day, many users have rightfully raised the concern of how to protect their data privacy. In response, laws such as the General Data Protection Regulation (GDPR), by the European Union (EU), and the California Consumer Privacy Act (CCPA) have been enacted to protect users' privacy. Common requirements call for providing access to or deleting all records in all ever-collected data given a specific user's search key(s) in a timely manner. The size of collected data and the volume of requests for search make enforcing GDPR and CCPA highly inefficient if not resourcefully infeasible. This use case examines how Delta Lake and Delta Engine can be used to help with compliance use cases, which are becoming more prevalent across the globe.

Relevant for

Any organization with consumer data with a particular focus on marketing and information security roles.

Challenges

Scale of data

When you have large amounts (petabyte-scale) of data in the cloud, user data can be stored and distributed across multiple data sets and locations.

Expense

Point or ad hoc queries to find data for specific users are expensive because they often require full table scans. Taking a brute force approach to GDPR/CCPA compliance can result in multiple jobs operating over different tables, leading to weeks of engineering and operational effort.

Feature tool set

Data lakes do not support the ability to perform row-level "delete" or "update" operations natively, which means that you must rewrite partitions of data. A typical data lake does not provide ACID transactional capabilities or efficient methods to find relevant data. Moreover, read/write consistency is also a concern, as read data is not protected from material impacts.

Data hygiene

Data hygiene in the data lake is challenging, given that data lakes by design support availability and partition tolerance with eventual consistency. Enforceable and rigorous practices and standards are required to ensure clean data.

databricks

Value with Databricks

With Databricks Lakehouse for Media & Entertainment, organizations can leverage Delta Lake as a transactional layer that provides structured data management on top of your data lake. It can dramatically simplify and speed up your ability to locate and remove personal information. Typical value levers include:

- **Improved query performance**

  Databricks helps maintain a high-performance pipeline that would support the cleanup of 5TB of data within 10 minutes.

- **Enhanced features**

  These include ACID capabilities overlaid on the data lake to prevent readers from being negatively affected when delete or update operations are carried out on the data lake.

- **Reduced cost**

  Processing costs and time to value are significantly reduced based on the automation that Databricks delivers for this use case.

Solution overview

Delta Lake is a very effective tool for addressing these GDPR and CCPA compliance requirements because its structured data management system adds transactional capabilities to your data lake. Delta Lake's well-organized, well-sized and well-indexed stats-enabled data sets enable quick and easy search, modification and cleanup of your data using standard SQL DML statements such as DELETE, UPDATE and MERGE INTO. Typical use case data sources include name, address, phone number, email address, IP address, device ID, lat/long, user ID and device ID.

How to get started

For more information on using Databricks Lakehouse for Media & Entertainment for GDPR and CCPA compliance, check out the following resources:

Best Practices: GDPR and CCPA Compliance Using Delta Lake

This article describes how Databricks Lakehouse for Media & Entertainment helps companies manage General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) compliance for your data lake. Because Delta Lake adds a transactional layer that provides structured data management on top of your data lake, it can dramatically simplify and speed up your ability to locate and remove personal information (also known as "personal data") in response to consumer GDPR or CCPA requests.

Make Your Data Lake CCPA Compliant With a Unified Approach to Data and Analytics

This how-to blog post reviews some of the functionality to enable within Delta Lake for a CCPA compliance use case on responding to a data subject request.

Case Study

Adobe™

CUSTOMER STORY ›

Adobe enforces GDPR and CCPA standards within the Adobe Experience Cloud and has realized a 10x reduction in search time when responding to a data subject request.

databricks

USE CASE: **Campaign Performance**

Overview

Modern marketers generally run digital campaigns across numerous digital platforms in addition to offline channels like television, radio, print and outdoor. Being able to track the performance of campaigns and aggregate this data to understand your most effective advertising channel is critical when building out your advertising strategy and understanding what level of return you are getting on your advertising spend. Capturing performance data can be a challenge, particularly in digital channels where impressions, clicks, conversions and page visits are all dynamic data sources that are tracked. This use case examines how to build real-time pipelines to understand advertising campaign performance across both digital and linear channels.

Relevant for

Agencies, brands and marketing organizations looking to understand a full view of both online and offline advertising campaign performance .

Challenges

Scale
Typical demand-side platforms might have to sift through hundreds of billions of advertising opportunities across dozens of ad-serving platforms.

Speed
Clients demand real-time insight into campaigns, so streaming pipelines are critical to meeting business-driven SLAs.

Cost and efficiency
Infrastructure native tooling generally lacks the performance and reliability benchmarks required for this use case, which often drives additional costs for processing.

Value with Databricks

Databricks Lakehouse for Media & Entertainment helps both publishers and marketers deliver a more performant pipeline for this real-time insights use case. Typical benchmarks from an ad tech demand-side platform (DSP) include:

• **Performance**
Databricks reliable and performant processing engine has reduced query times from 24 hours to 3 hours for large jobs.

• **Scale**
Able to ingest real-time data from 40 ad exchanges and execute against 200 billion ad opportunities daily.

databricks

Solution overview

Databricks Lakehouse for Media & Entertainment is an optimal solution for running this kind of real-time workload, as it brings together all of the workflows in one place. The Databricks notebook-based UI suits this workflow well in addition to being able to extend functionality for ETL, logic, logging and extra monitoring needs. These classes can be imported into a notebook, and the Databricks jobs scheduler can be used to configure the details of the cluster that this workflow runs on to run the job itself. The convenience of having one place to manage these use cases speeds up the development and production process compared with other solutions in the market. Typical use case data sources include:

- **Channel, channel type, marketing objective (e.g., awareness, conversion)**

- **Metrics:** impressions, clicks, CTR, media spend, revenue (e.g., subscription, sales, ads)

How to get started

To see how MediaMath architected a real-time campaign performance engine with Databricks Lakehouse for Media & Entertainment, check out the following customer presentation:

Extreme-Scale Ad Tech Using Spark and Databricks at MediaMath

MediaMath is a leading ad tech platform that responds to over 200 billion ad opportunities daily, and leverages massive amounts of data to power smarter digital marketing. They use Databricks heavily both in production and R&D to develop innovative, proprietary and scalable solutions for predicting conversion probability given an ad impression. This allows them to measure causal effectiveness of advertising, run simulations to understand the impact of cookie refreshes and other phenomena on ad effectiveness metrics, and find device IDs belonging to the same user based on possibly noisy external deterministic information.

Case Study

CUSTOMER STORY >

Learn how MediaMath tracks real-time campaign performance across 4 billion events a day across 40 ad-serving platforms.

databricks

>

USE CASE: **Yield and Pricing**

Overview

Yield management is a complex technical process that involves being able to foretell advertising demand and manage inventory (also known as stewardship). Being able to forecast inventory correctly and manage the sales pipeline effectively becomes two levers for an effective yield management strategy. Publishers can maximize yield only by allocating inventory in the areas where the demand seems to be highest. The only way for publishers to effectively deploy their products is by having complete visibility into their revenue across all channels.

The means of increasing revenue in the course of the execution of inventory control and pricing strategy is advertising yield management. This use case examines how Databricks Lakehouse for Media & Entertainment is being used to drive better inventory management and revenue per unit outcomes.

Relevant for

Advertising publishers including broadcasters, print and digital media publishers, out of home, and other organizations selling advertising.

Challenges

Ability to handle streaming data
Streaming sources like Nielsen and Omniture need to be accessed in real time to understand performance and pricing.

Limited access to data
Advertising performance and ratings data need to be accessible alongside consumer insights and segmentation, which are the foundation for pricing decisions.

Lack of data science/ML
A scalable data science platform needs to predict audience sizes for linear and digital channels to drive accuracy in inventory pricing.

Value with Databricks

Typical value benchmarks from a broadcaster leveraging Databricks Lakehouse for Media & Entertainment for this use case include:

- **Cost reduction**
  40% monthly cost reduction of native infrastructure solution.

- **User productivity**
  30% increase in productivity, allowing data engineers and data scientists to focus on more value–added projects over infrastructure maintenance.

- **Time to insights**
  Improves insights and relationships with data across viewers, brands and content.
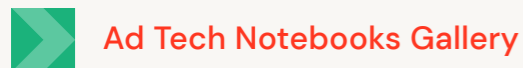
databricks

Solution overview

Running this use case on Databricks Lakehouse for Media & Entertainment helps drive down runtime by 50% over using native infrastructure solutions and also improves resource monitoring and observability. Delta Lake allows for improved performance and reliability, including the ability to do time travel and to use MLflow for collaborative notebooks and multi-language support. Typical use case data sources include:

- Campaign name, campaign start date, campaign end date, franchise, demo, "competing" campaigns, ad server priority, frequency caps
- Guaranteed impressions, impressions, eCPM

How to get started

Databricks has a number of assets and Solution Accelerators to assist with advertising use cases. Some of these assets include:

> **Ad Tech Notebooks Gallery**

Data analysis and data science are at the core of optimizing advertising campaigns. This gallery showcases some of the possibilities through notebooks that can easily be imported into your own Databricks environment.

> **Solution Accelerators for Media & Entertainment**
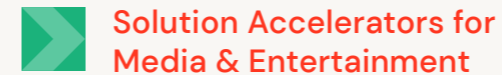
Based on best practices from our work with the leading brands, we've developed Solution Accelerators for common analytics and machine learning use cases to save weeks or months of development time for your data engineers and data scientists.

Case Study

Paramount   **CUSTOMER STORY >**

Hear how Paramount is leveraging Databricks Lakehouse for Media & Entertainment across a number of audience experience- and advertising-based use cases.

databricks

USE CASE: # Inventory Forecasting

Overview

Programmatic advertising has grown dramatically in recent years as media buyers are looking for more efficiency than the traditional IO process in traditional digital advertising and the ability to make campaign optimizations in real time. As sellers and media owners move toward more programmatic approaches, they need to ensure that they have enough inventory to fulfill customer requests. Armed with forecasting insights, media owners can intelligently plan for the future and strategize how to package and sell their inventory. This use case examines how to establish real-time pipelines for inventory forecasting so sellers can iterate, propose and sell ad campaigns faster.

Relevant for

Publishers selling advertising inventory, ad networks, media buyers executing ad buys, agencies.

Challenges

Data scale
Ad networks are processing tens of billions of events a day.

Ability to drive real-time insights
Ad inventory is very dynamic, so the ability to have real-time pipelines is a requirement for this use case.

Cost
Many customers struggle with leveraging native infrastructure to build reliable and performant pipelines for real-time insights.

Value with Databricks

Databricks Lakehouse for Media & Entertainment has helped a publisher ad network achieve the following performance gains:

- **Infrastructure cost improvement**
  2x lower infrastructure costs.

- **Scale**
  35 billion+ events per day processed for analytics.

- **Performance**
  5x faster data pipelines.

databricks

Solution overview

Databricks Lakehouse for Media & Entertainment is the foundation for this solution, as it leverages Delta Lake as an open source storage layer that brings ACID transactions to Apache Spark for streaming and batch data sets. Data is landed in an ingestions table (Bronze) and moved to a refined table (Silver), and a feature store (Gold) is added to allow for downstream analytics and machine learning use cases, all within your existing data lake. Typical use case data sources include:

- Date, property, franchise, advertiser, demo, ad unit, device type, geo, frequency caps
- Unique devices, impressions

How to get started

To see how GumGum architected a real-time inventory forecasting with Databricks Lakehouse for Media & Entertainment, check out the following customer presentation:

Real-Time Forecasting at Scale Using Databricks and Delta

GumGum receives around 30 billion programmatic inventory impressions amounting to 25TB of data each day. Inventory impression is the real estate to show potential ads on a publisher page. By generating a near real-time inventory forecast based on campaign-specific targeting rules, GumGum enables account managers to set up successful future campaigns.

Case Study

gumgum

CUSTOMER STORY >

With a unified approach to data and machine learning powered by Databricks Lakehouse for Media & Entertainment, GumGum is able to accelerate response times, reduce operational costs and deliver real-time inventory forecasting across their publisher network.

databricks

USE CASE: # Multi–Touch Attribution

Overview

Today's buyer journeys span many devices and touchpoints before resulting in a conversion. In order to optimize campaigns and create more customized consumer experiences, marketers need to understand which touchpoints and messages a consumer came in contact with that resulted in positive action. Multi–touch attribution is a method of marketing measurement that looks at the impact of each consumer touchpoint in driving a conversion, thereby determining the value of that specific touchpoint relative to the marketing budget. The goal is to understand where to allocate your marketing spend. When marketers can understand the role that certain touchpoints played in a conversion, they can more effectively devote funds to similar touchpoints and divert funds from ineffective touchpoints. This use case looks at how to build an MTA model to understand marketing effectiveness.

Relevant for

Agencies and marketing organizations looking to understand the effectiveness of their marketing efforts relative to marketing spend.

Challenges

Lack of offline metrics
MTA is largely used for campaigns that employ digital marketing platforms, as it measures consumer actions (e.g., clicks). This makes it challenging for these models to incorporate offline data, such as exposure to a TV or print ad.

Data wrangling
The volume of data and complexity of the building and training models pose a challenge, as marketing analysts tend to spend more time aggregating than deriving meaningful insights from it.

Lack of visibility into external factors
Unlike media mix modeling, which looks at aggregate data, MTA looks at user-level insights. Without incorporating aggregate information, marketers do not have visibility into external trends that might affect marketing efforts and conversions, such as seasonality.

Value with Databricks

With Databricks Lakehouse for Media & Entertainment as the foundation for their multi–touch attribution solution to understand marketing spend across channels, sample value metrics from a media agency include:

- **Infrastructure cost savings**
  22% lower operational cost of running their data pipelines year over year.

- **Increased data engineering/data science collaboration**
  30% productivity improvement across data teams.

databricks

Solution overview

Databricks Lakehouse for Media & Entertainment enables organizations to leverage Delta Lake to combine various sources of data in a timely and efficient manner, including batch and streaming data sources that are needed to measure both online and offline campaign performance. MLflow is also a critical component of this use case because data science techniques are applied on top of the campaign performance data to make recommendations to remediate ineffective campaigns. Typical use case data sources include:

- User/device ID, channel, channel type, creative type, creative name, campaign start date, campaign end date
- Revenue, conversions, marginal contribution, current spend allocation, recommended spend allocation, estimated lift from reallocating spend

How to get started

> **Multi-Touch Attribution Acclerator**
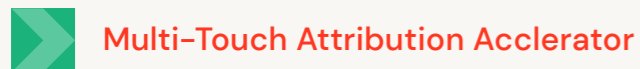
Attracting high-quality users as cost effectively as possible is the goal of every marketing organization spending money on advertising. But with dollars dispersed across a wide array of campaigns, channels and creatives, it can be hard to understand the most effective ways to track return on ad spend.

The multi-touch attribution accelerator enables data teams with fine-grained insights and real-time dashboards to help you optimize your media mix and easily integrate new marketing channels into your campaign performance models.

> **Sales Forecasting and Advertising Attribution Dashboard Solution Accelerator**
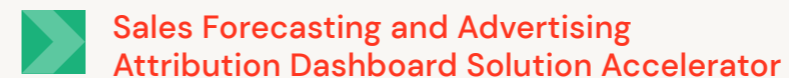
Whether you're an ad agency or an in-house marketing analytics team, this Solution Accelerator allows you to easily plug in sales data, ad engagement data and geodata from a variety of historical and current sources to see how these drive sales at a local level. With this solution, you can also attribute digital marketing efforts at the aggregate trend level without cookie/device ID tracking and mapping, which has become a bigger concern with the news of Apple deprecating IDFA.

Case Study

PUBLICIS GROUPE

CUSTOMER STORY >

In addition to scaling an omnichannel personalization platform with Databricks Lakehouse for Media & Entertainment, Publicis Epsilon has also scaled an attribution model from the platform in order to understand the effectiveness of its personalization efforts.

databricks

USE CASE: # Segmentation

Overview

Segmentation is a process of dividing a heterogeneous market into relatively more homogenous segments based on certain parameters like geographic, demographic, psychographic and behavioral. It is the activity of dividing a broad consumer or business market, normally consisting of existing and potential customers, into subgroups of consumers (known as segments) based on some type of shared characteristics.

Relevant for

Marketers looking to drive more precision in their targeting efforts, publishers that want to add more attributes to audience data to increase its requisite value.

Challenges

Lack of machine learning capability
Customer segmentation today involves basic analytics on a narrow set of variables such as age, gender and income.

Limited data
Narrow inputs often lead media companies to view each cohort as a homogenized group, limiting everything from personalization to ad targeting.

Poor insights
Inability to augment basic customer segments for better insights into the churn, upsell and other lifecycle behaviors.

Value with Databricks

With Databricks Lakehouse for Media & Entertainment as the foundation for their data analytics and machine learning efforts, a sample publisher realized the following in improvements:

- **Processing time**
  60% reduction in processing times due to faster ETL pipelines.

- **Cost**
  50% reduction in IT operational costs.

- **Productivity**
  Faster time-to-insight enabled them to experience significant growth in business.

Solution overview

Databricks Lakehouse for Media & Entertainment enables organizations to leverage Delta Lake to combine various sources of data — including transactions, customer demographics across products and services with digital journeys, and marketing analytics — in a timely and efficient manner to bring a 360 view of customer interaction. With millions of customers, large organizations are able to leverage AI to extract similar customer patterns and segment customers no longer based on who they are but how they will engage, bringing a more inclusive approach to segmentation. Typical use case data sources include:

- Marketing campaigns: engagement, winback

- Device ID, email address, "fan of ____ ," subscription length, subscription expiration date, subscription status, date of last use, assigned audience segment

- The propensity to churn, engagement score, LTV, video views, gameplays

databricks

How to get started

Behavioral Segmentation Solution Accelerator

In this series of notebooks, we demonstrate how to create advanced customer segments to drive better purchasing predictions based on behaviors. Using sales data, campaigns and promotions systems, this solution helps derive a number of features that capture the behavior of various households, which allows users to build useful customer clusters to target with different promos and offers.

Reach Frequency Monetary Amount (RFM) Solution Accelerator

In this notebook, we demonstrate a simple application of RFM segmentation based on customer purchase histories. We modernize the technique a bit, applying some basic cluster analysis to identify groups of customers who are similar in their RFM scores, which we may target in different manners based on their different potential.

Case Study

CONDÉ NAST

CUSTOMER STORY >

Condé Nast invests heavily in products to improve and enhance the audience's experience. One such product solution is Spire, Condé Nast's service for user segmentation and targeted advertising for over 100 million users. Spire consists of thousands of models, many of which require individual scheduling and optimization. Learn how Condé Nast navigates the complexities of building large-scale machine learning pipelines within Spire leveraging Databricks Lakehouse for Media & Entertainment.

databricks

USE CASE: # Sentiment Analysis

Overview

Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation and understand customers. Since customers express their thoughts and feelings more openly than ever before — particularly on social channels not owned by brands — sentiment analysis is becoming an essential tool for monitoring and understanding that sentiment. Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated so that they can tailor products and services to meet their customers' needs. This use case examines the best practices in using NLP techniques to determine end-user sentiment.

Relevant for

Product teams looking to understand the sentiment on new product features, marketers trying to understand brand/product reputation and service teams attempting to be proactive in finding issues across social channels.

Challenges

Iterative query development
Need to build ML models that are capable of providing personalized in-app offers to millions of monthly users and match them against user profiles to ascertain sentiment.

Slow data pulls
Manually monitoring petabytes of streaming network data across tens of thousands of city and ISP configurations is nearly impossible, making it hard to proactively pinpoint network issues that adversely affect user experience and sentiment.

Disjointed infrastructure
Moving data across disjointed systems and data analytics tools hinders team agility and collaboration, particularly when scraping sentiment data from rigid, social platforms.

Solution overview

Databricks Lakehouse for Media & Entertainment streamlines analytics workflows across cross-functional teams with a single platform for querying, debugging and exploring streaming and batch data as well as building and deploying ML models. The ability to handle streaming data from social platforms allows for sentiment or service issues to be known and addressed in near real-time. A simplified data management platform allows organizations to fully automate job scheduling, monitoring and cluster management without human intervention. Typical use case data sources include:

- Customer ID (optional), customer feedback, feature, sentiment and social

databricks

How to get started

> ▶ **Tutorial: Sentiment Analysis on Streaming Data Using Azure Databricks**

In this tutorial, you learn how to run sentiment analysis on a stream of data using Databricks Lakehouse for Media & Entertainment on Azure in near real-time. You set up a data ingestion system using Azure Event Hubs. Next, you consume the messages from Event Hubs into Databricks using the Spark Event Hubs connector. Finally, you use Cognitive Service APIs to run sentiment analysis on the streamed data. By the end of this tutorial, you will have streamed tweets from Twitter that have the term "Azure" in them and run sentiment analysis on the tweets.

> ▶ **The Pursuit of Happiness: Building a Scalable Pipeline Using Apache Spark and NLP to Measure Customer Service Quality**

In this video, the presenters explore how to expand NLP/sentiment analysis to investigate the intense interactions that can occur between humans and humans, and between humans and robots.

Case Study

JAM CITY

CUSTOMER STORY ›

Jam City monitors in-game player sentiment to understand gamer satisfaction and leverage the data into their product development lifecycle.

databricks
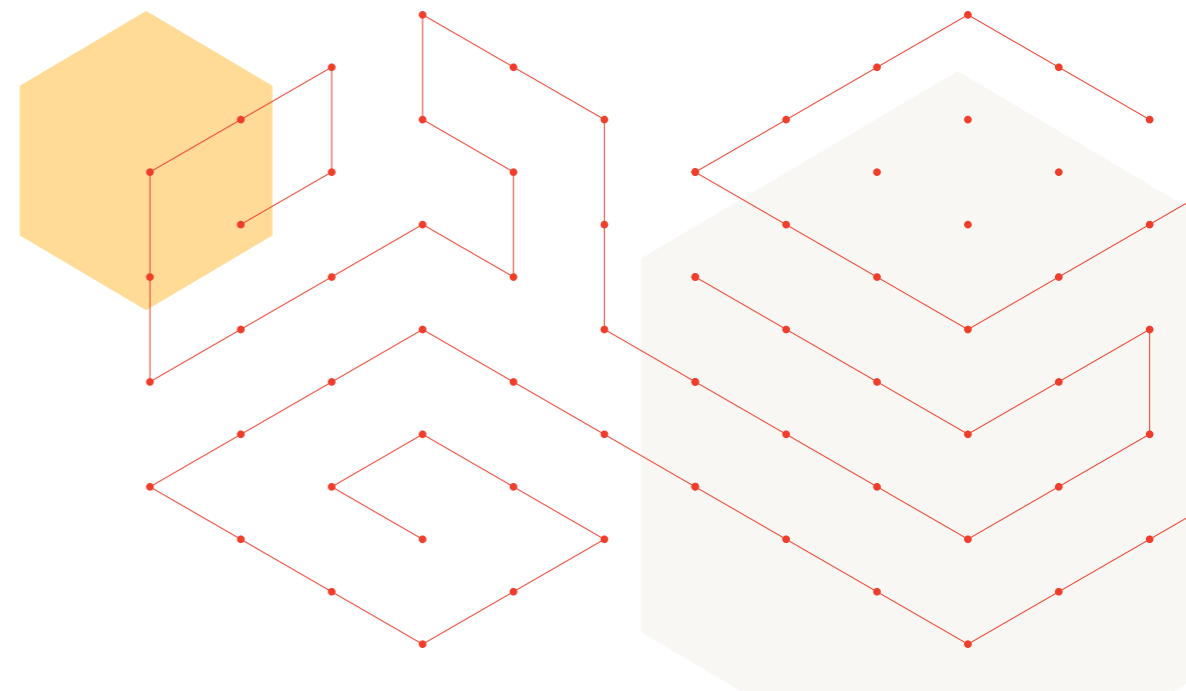
USE CASE: # Toxicity Detection in Gaming

Overview

Across various forms of online gaming, players continuously interact with one another in real time to either coordinate or compete as they move toward a common goal. This interactivity is integral to gameplay dynamics, but at the same time, it's a prime opening for toxic behavior. Such behavior manifests in many forms, but the most severe toxicity tends to happen through in-game chat (text/audio), which happens to be a perfect fit for addressing with ML/AI. In this use case, we demonstrate how to detect toxicity in in-game chat using natural language processing.

Relevant for

Product teams and community managers looking to use automation and more efficient moderation to better cultivate a healthy and inclusive environment for gamers.

Challenges

Ability to handle streaming data
Large volumes of unstructured data coming in from multiple sources, in multiple file formats and in near real-time.

Natural language processing
Using natural language processing at scale to detect toxicity in in-game chat across multiple languages and multiple games.

Community moderation
Helping internal community managers sift through in-game chat data more efficiently to intervene and curb toxic behavior.

Solution overview

Executing on this use case requires support for processing unstructured data in near real-time, performing inference using natural language processing and easily integrating with downstream systems such as community moderation tools. With the Databricks Lakehouse for Media & Entertainment, this can be achieved within a single platform, making it a perfect fit for this use case. Further, all of this data can be easily combined with activity data to understand the impact that toxicity has on player churn and engagement.

databricks

How to get started

> **Solution Accelerator:
> Toxicity Detection in Gaming**

In this Solution Accelerator, we walk through the steps required for detecting toxicity in in-game chat in near real-time. This includes streaming data into Delta Lake, using Spark NLP to classify comments as belonging to various forms of toxicity, and performing inference on both batch and streaming data.

> **How to Use ML to Detect Bullying Behavior in
> Gaming Communities — The Databricks Blog**
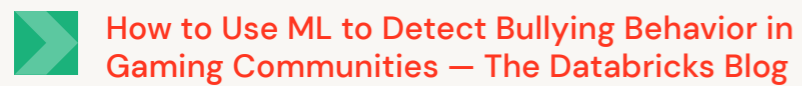
Learn how to use Labelbox/Databricks for model-assisted labeling so that you can quickly annotate your own data for modeling training.

Case Study

SEGA®

CUSTOMER STORY >

SEGA monitors in-game player sentiment and looks for toxic behavior among players which is a key indicator of churn and customer satisfaction.
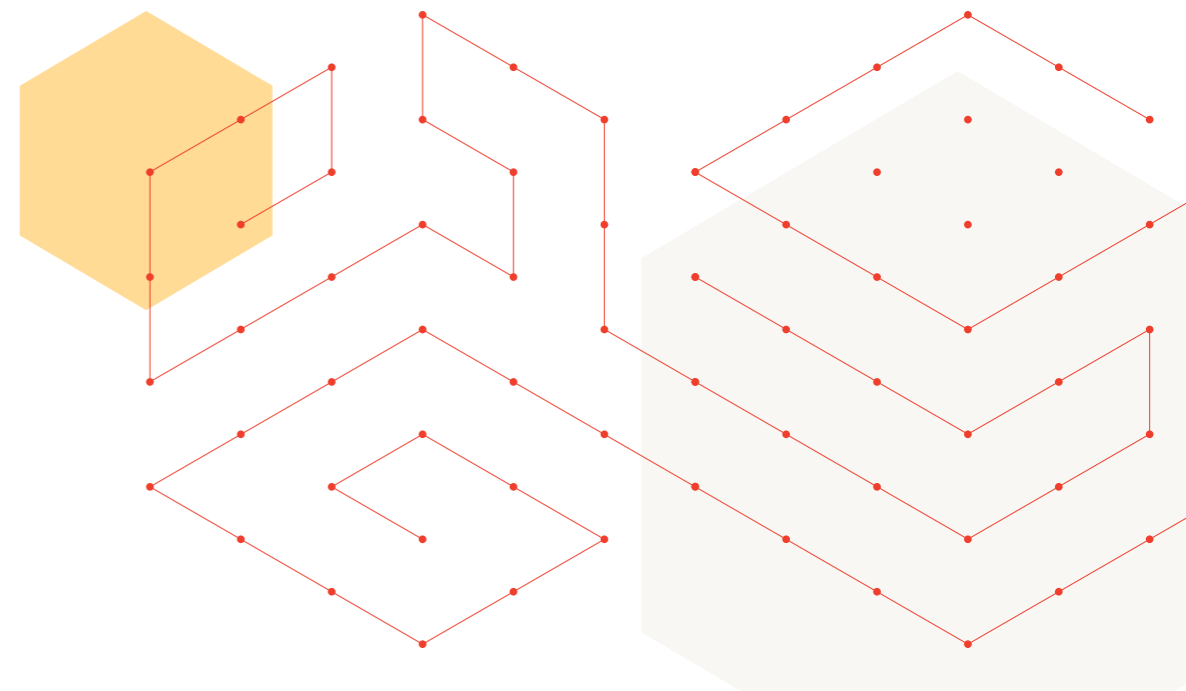
databricks

Conclusion

Today, data is at the core of every innovation in the media and entertainment industry. Databricks Lakehouse for Media & Entertainment enables companies across video, gaming, music and more to harness the power of data and analytics to meet the needs of a real-time, on-demand world and provide superior customer engagement to be in a position to win.

Get started with a free trial of Databricks Lakehouse for Media & Entertainment and begin building better data applications today.

START YOUR FREE TRIAL

Contact us for a personalized demo
databricks.com/contact

databricks