



WHITE PAPER

The key to scalable insights in digital data operations

Access to growing amounts of data should be an asset, not a liability. Organizations have an unprecedented opportunity to grow their business by taking a scalable, automated approach to managing their data. This is what we call data operations (DataOps).

Connect with us at sales@switchboard-software.com

The key to scalable insights in digital data operations



Digital channels have undergone a fundamental shift where data now drives our understanding of supply inventory, customer demand, customer needs, and profitability. The sheer volume of data coming each day from social, affiliate ecommerce, newsletter subscribers, subscription paywalls, acquisition funnels and third-party cookies makes the task of extracting meaningful insight more complex than ever before.

This, in turn, has put a premium on new tools and techniques to measure performance. Yet data complexity, volume, and engineering challenges continue to block even the most advanced business teams. The emerging science of DataOps has been evolving to meet this challenge, and organizations now stand at the crossroads of data and efficiency. Either you master your data so your business can evolve based on quantified insight, or your data will keep your teams in an ever-lengthening cycle of manual reporting.

This paper was written by Switchboard's co-founders, who helped to launch Google BigQuery in 2012. With our extended team at Switchboard, we bring a wealth of experience building agile and scalable data platforms for leading digital organizations.

This paper was created to help CROs, CMOs and CDOs learn:

- ✓ How DataOps can unify and automate the measurement of monetization, inventory, audience reach and content performance in real time
- ✓ Which data sources and formats organizations must tie together to create timely and accurate business KPIs
- ✓ The best approaches to help you consolidate disparate data in a scalable manner, and the strengths and weaknesses of these approaches

Access to growing amounts of data should be an asset, not a liability. Digital organizations have an unprecedented opportunity to grow their business by taking a scalable, automated approach to managing their data. We hope you find this paper helpful and we welcome your thoughts and feedback at sales@switchboard-software.com.

What is Data Operations?



DataOps is both a process and an approach for empowering business teams to extract value from diverse datasets in order to deliver real-time, business-specific insights. When it comes to new ways of managing data, it's good to set out all of your requirements, before you identify and invest in a solution, to make sure it will yield your intended business results. With DataOps for organizations, you can accomplish this by evaluating the following core tenets:

- ✓ Data can be easily configured once and automated.
- ✓ Data is managed by the same repeatable process, regardless of scale.
- ✓ Data can be monitored 24/7, and any disruption handled without data loss or corruption.
- ✓ Data provenance can be audited so that it can be used for mission-critical decisions.

New data operations challenges with digital data

The data landscape is becoming increasingly complex. Revenue is no longer driven by a single Google Ad Manager (GAM) instance to be managed by a lean AdOps team. Programmatic sales, mobile, affiliate ecommerce, social media, newsletter subscribers, subscription paywalls and acquisition funnels all provide new revenue channels, but they fragment your view of inventory, revenue and yield. Advertisers may be willing to pay a premium to reach specific audiences, but only if they can be assured that those demographic groups are viewing their campaigns.

Revenue and AdOps teams are often burdened with requests to pull raw data manually and painstakingly stitch it together into one-off reports. The intense pressure caused by the need for operational visibility creates a dangerous stumbling block. While at Google, we worked with some of the world's largest organizations and ad agencies to enable real-time insights from a rich mix of media data. We saw firsthand how difficult these data challenges can be.

Forward-thinking organizations know they need timely data insights to effectively grow customer loyalty, satisfy advertisers and optimize yield. Organizations that do not use data effectively today risk being outflanked by more nimble competitors tomorrow, as they drown in a sea of disparate data. Unless they have the right strategy and the right tools.

A scalable DataOps solution delivers continuous visibility into critical KPIs like Sell-Through Rates, Campaign Delivery, and User Growth. These insights help you become aware of emerging trends, rapidly understand their causes, and use those insights to create new opportunities for growth and profitability. So, how can you use DataOps to empower your digital business?

The path to real-time insights

DataOps is an approach that combines processes for collaborative data management, tools for automation and monitoring, and a scalable architecture to ensure that data growth is an asset, not a liability.

There are four fundamental steps to realizing the benefits of DataOps:

01

Identify

KPIs that will help you measure and improve performance

02

Normalize

raw data into foundational data

03

Transform

foundational data to create distinct KPIs

04

Automate

using the right tools for DataOps and real-time reporting

01

Identify KPIs that will help you measure and improve performance

A DataOps approach aims to provide value by enabling organizational data collaboration. Unfortunately, in their rush to set up metrics, organizations will often base their choice of measurements on whatever data is available or is easiest to present. This inevitably results in arduous, manual work when it comes to providing higher-level insights. Developing a DataOps strategy starts not with technology, but with organizational communication. Your initial goal should be to recognize, at a high level, which specific metrics you want to measure. At the outset, it's essential that you target questions that will yield the insights you need to drive key decisions.

Here are some common business questions teams should ask:

- ✓ Where are we likely to have sell-out or unfilled inventory over the next 30 days?
- ✓ How are monetization rates changing across our verticals over time?
- ✓ Which programmatic deals bring us the most revenue and highest eCPM?
- ✓ What is the ROAS, CAC and LTV of our customers?
- ✓ What are our most valuable audiences in terms of viewability and sell-through?
- ✓ How can we best optimize our sales between programmatic versus direct?

01

Here are some common examples of core KPIs that can be enabled by DataOps to answer those questions:

Metric	Definition
Sell-through Rate	How much of the available inventory has been monetized (ideally broken down by specific verticals or ad units)?
Impression Delivery	Have impressions been delivered to campaigns as required, in the necessary numbers and from the appropriate inventory sources, such as Google Ad Manager (GAM), YouTube, Programmatic?
Blended eCPM	What is your eCPM for individual groups of ad products or verticals?
Programmatic Deals	Which programmatic sources are producing the most revenue and at the highest efficiency? What is the current and historical performance of different programmatic deal types and partners?
Revenue per User	How much ad revenue are you generating per visitor? What is their Life Time Value?
Revenue Capacity	What is the total addressable revenue opportunity across all inventory and demand sources? Where are the biggest opportunities to increase revenue and profit?
Inventory Forecast and Availability	How much inventory will you have to sell looking forward (ideally broken down by vertical, ad unit, audience, viewability)?

While it's possible to generate these metrics with manual techniques, you risk burning valuable talent on repetitive and error-prone tasks: logging into numerous tools and user interfaces, and downloading, scrubbing and wrestling unstructured CSV files into rudimentary spreadsheets. DataOps automates these tedious processes, so questions that used to take weeks to answer can now be handled in minutes.

02

Normalize raw data into foundational data

Once your data collaboration goals and key metrics are established, the technical work of producing value from your data begins. Various teams in an organization require different combinations of metrics. But raw data accessed directly from APIs or log files rarely provides a format useful for collaborative analysis. An emerging pattern to address this issue is to continuously and automatically refine raw data streams into a format useful to derive KPIs. This is what we call 'foundational data'.

Foundational data: The basis for true KPIs

Just as a jet aircraft requires highly-refined aviation fuel to achieve its full potential, similarly the KPIs that drive your business decisions require data that is high quality, well-understood and highly reliable. Unfortunately, raw data that comes from vendors and third parties can be anything but. No two APIs are alike, so data must be cleaned, typed and sometimes enriched with match tables in order to be useful. Sometimes, data formats change, while connectivity or vendor hiccups present the ever-present risk of data loss or data corruption.

Building meaningful KPIs directly from such data is impossible without taking on an enormous amount of complexity. Systems built upon such data start out brittle, eventually collapsing under their own weight. With such a large gap between raw data and KPIs, experienced teams invest in an intermediate concept that is critical for success — foundational data. The key idea is to take each source and normalize the data into standardized and canonical versions. Ideally, those that can be easily combined with other, similarly-refined data sources.

Common data sources publishers need to normalize

Within most organizations, the list of data sources you need to master is clear. However, each incremental data source adds new complexities. By understanding the distinctive properties and challenges presented by each source, you'll be able to make a more informed tool selection based on the unique profile of your business.

Disparate data sources, formats, and integration challenges will sap any budget, so how can organizations protect against mounting costs?

02

Rather than attempt to on-board every single data source for its own sake, try to understand the data characteristics of your business today, and where it will be tomorrow. This will help you understand how your raw data can evolve to become the foundation for the specific metrics you need to succeed.

Creating foundational data – an example using GAM

Let's say you're most interested in creating foundational GAM data — because as your primary ad server, it can provide a rich view of how certain display and video inventory is being delivered. Start with the GAM API. For brevity, we'll assume you're already familiar with its quirks and limitations.

The first step is to determine the appropriate queries and granularity of data required. An important consideration is identifying the dimensions you really need as there are quota limits. Next is to use a script or a tool to invoke the API, and extract and store the query result. It's important to do this with 100% consistency so that query results maintain the same schema. Each row in the query result needs to be type-checked, i.e., numeric values must be cast into integers or floats if they are to have any value for calculations. Dimensions must be normalized in order to avoid textual inconsistencies (as a result of occasional human input error) that can also throw off calculations. Finally, the query result needs to be written either to a file or, preferably, to a data warehouse, so that it can be consolidated for queryability.

Additional considerations include how to extract key-values so that the business attributes captured in custom dimensions can be extracted for analysis, as well as backfilling. Lastly, consider if you need Data Transfer, the event-level server logs that can provide the finest possible granularity and level of insights.

The steps above are the abbreviated set of tasks involved in creating foundational data for a selected data source. The processes involve numerous data cleansing tasks (file encoding, to name just one example) that go beyond the scope of this paper.

03

Transform foundational data to create distinct KPIs

Going from raw data to KPIs requires a few steps. Once you've established a clear list of KPIs to measure performance, and marshaled the raw data from various systems and vendors, how do you connect these two concepts? In many cases, crucial KPIs must be derived from different data sources with unique schemas and formats.

To fully realize how the overall value of your digital data is greater than the sum of your individual channels, each data source must be prepared based on its unique properties. DataOps automates many styles of data preparation into a single process that transforms raw ingredients from any data source to create foundational data. This powerful type of data can then be shaped into KPIs that power better business decision-making. The diagram below illustrates the gap foundational data fills to help you move from data to intelligence.

The gap between raw data and KPIs



03

Even a single KPI can require blending multiple sources

To understand this better, let's look at a KPI from Step 1 as an example: Sell-Through Rates. Let's say Revenue Operations needs to pinpoint where the highest and lowest Sell-Through Rates occur across multiple properties. These metrics could be derived by analyzing delivery logs from the ad server. But what if the client demands only impressions that were viewable? Deriving the answer requires combining delivery data with impression data from the viewability measurement system.

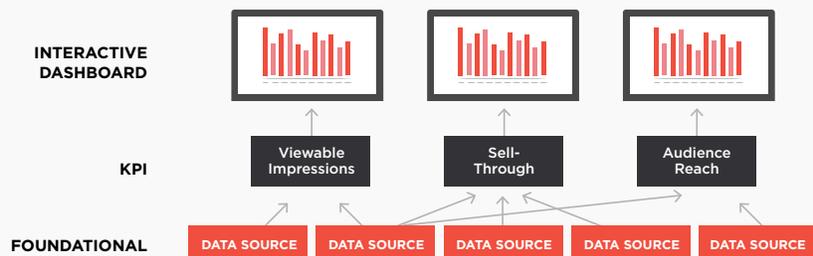
Foundational data: A scalable digital data hierarchy

The kind of data blending described above quickly gets unwieldy without a structured approach. Foundational data is a layer of data derived from raw data. The focus is on modeling canonical "base metrics", such as Impression Delivery, Viewability and Revenue. These base metrics can then be used to derive more complex KPIs, for example, Viewable Sell-Through or eCPM.

There are huge benefits to the foundational data approach. It enables you to establish common data standards between multiple teams across the organization. It also encourages efficient reuse of data for different analyses and reports, meaning your entire organization starts to benefit through more timely and accurate communication, reduced errors, and ultimately, better decision-making.

You may have already created a data science team to establish foundational data. While data scientists can help to combine raw data and analyze the KPIs, even the most skilled data science teams lose valuable time deriving a KPI from scratch. When digital sales and executive leadership come to depend on this data, they will inevitably demand fast and reliable reporting of these metrics—and to do that, you'll need to make sure you pick the right tools.

Foundational data provides the basis for a scalable data hierarchy



04

Automate using the right tools for DataOps and real-time reporting

The value of the metrics from the last section of this paper should be unassailable to digital organizations, which begs the question: “Why aren’t we doing this now?”. The most common reason is operational capability. KPIs and foundational data can only be useful if the integrity and dependability of the underlying data is unquestionable. Such an endeavor is a complex technology problem, requiring real-time integration of heterogeneous data streams on top of a rock-solid and highly scalable operations platform. To address these issues, a DataOps approach takes advantage of scalable technology, including cloud data warehouses, as well as software that monitors performance and validity at every step of the data pipeline.

Harnessing raw data with efficiency

With a prioritized list of data sources and an understanding of how each data source and KPI will be handled, next you will want to weigh carefully several important considerations to ensure data sources are managed in a cost-efficient and scalable manner. For example:

Monitoring: How will third-party API uptime be monitored to ensure reliable delivery?

Problem triage: Once we’re aware of a problem, how will we pinpoint if it’s coming from the APIs, or the data warehouse, or somewhere in between?

Data quality: If a segment of data fails to load, or some portion of the data was malformed, how will we know? And how will we recover?

Data synchronization: How will we ensure that KPIs that depend on multiple sources have up-to-date components?

Change management: What happens when an API or data format changes?

Data scale: How will we scale up our processing capacity to handle event-level data that grows to billions of rows per month?

Operational re-use: Can the capabilities developed for one set of data sources be applied to all of my APIs and file-based sources, so that teams can collaborate using a single approach?

04

This is not an organizational problem, this is a data problem that requires automation. So, as an organization, you can try to build an expensive and highly specialized team to write and maintain custom infrastructure, or hire highly-priced consultants to do the same thing. Neither approach will deliver a long-term solution that takes advantage of DataOps best practices in a cost-effective way.

Cloud data warehouses make consolidating data significantly easier

Consolidating data streams into one place requires tools that are always available, and scale affordably to handle growing amounts of data. Within the past few years, three commercial cloud-hosted solutions – Google’s BigQuery, Snowflake, and Amazon RedShift – have proven themselves best-in-class for this task. However, some organizations still invest in a “Do-It-Yourself” approach, using traditional IT tools, developing custom software, and staffing ops engineers to maintain on-premises systems.

Foundational data: The basis for digital data insights

Deriving business metrics from raw data in a bottom-up manner is one of the most powerful tools available to any organization's. This can only be achieved by transforming raw data into foundational data, an intermediate step between the unstructured and the insightful. Foundational data is the canonical representation of data in its purest form. As an example, consider impression delivery data from GAM. Pulling manual GAM reports may provide the data needed to understand campaign delivery for a specific set of advertisers or key-values. But if you want to look at a different slice of data or time-frame, you need a different report, which means someone on your team needs to manually configure certain dimensions, time-frames, geographies, and so on. This means you never have a "single source of truth" to slice and dice, explore, or connect to other data. In other words, you never have

the basic building blocks that consistently tell the same story in the same language. Now, what if it was possible to pull down a complete repository of your GAM data, for every Line Item, Advertiser, Campaign, and Ad Unit, and across all geographies, properties, and time-frames? Then clean it up so that names, IDs and key-values are consistent across all datasets? That valuable transformation of raw data is foundational data.

Ask Yourself:

"Is it more efficient for me to build a new internal team with expensive and highly specific expertise to manage data, or will my existing team deliver better business results by relying on proven technology from an established vendor instead?"



In working with one of our large international customers focused on financial news, it's not that they can't handle basic data—but that they realized it doesn't make sense to build a large internal team dedicated to basic data management when they could be investing the same resources in creating innovative advertising products. At the end of the day, the business objective is to maximize client results and increase CPMs.

— Ju-kay Kwek, Switchboard CEO

ROI in digital organizations

Leveraging previously hard-to-use data to increase the performance of organization operations is an exciting new stage in the evolution of digital media. By combining data expertise with the wealth of fine-grained data available from so many data sources, it's an exciting time to be in Sales, Marketing, Revenue Operations, AdOps or Data Science in this industry.

As data volumes and sources continue to soar, compelling ROI data from Switchboard customers is proving the value of DataOps for any data-driven enterprise. Here are summary ROI stats from two companies with whom we've been working:

Financial Times

- ✓ Stood up sell-through forecast in one week; CPH ops solution in three weeks
- ✓ Rolled out real-time forecasting available to all yield, planning, and ops teams, eliminating a manual bottleneck of four-person days per forecast
- ✓ Large-scale data automation and governance was delivered via a hosted platform, requiring no additional developer investment

Dotdash Meredith

- ✓ Two billion rows of daily data harnessed to drive ad revenue for brand partners
- ✓ Six months development and data preparation saved
- ✓ One source of truth for revenue reporting across 40 brands



Conclusion

Bringing the diversity and volume of data that exists within a digital organization into a single revenue-driving software system isn't easy—but the rewards are worth the effort. Once the operational foundation is in place to turn raw data into foundational data and KPIs, you can move rapidly towards advanced analytics such as working with event-level data. At that point, insights such as user behavior cohorts, programmatic bid-ask spreads, and fine-grained revenue forecasting will be within your grasp.

It's important to make an honest assessment of your company's skills and existing infrastructure to decide which approach will be right for you. When the time comes, use that same critical eye to gauge what kind of data consumption you expect in order to pick the right solution architecture for your business.

At Switchboard, we understand that the most difficult challenges when working with media data involves the engineering effort necessary to maintain flexible, scalable and operationally robust data flows. Using a tool like Switchboard will help keep valuable technical staff focused on helping make decisions instead of learning a complex new engineering competency.