

Content Distribution Analysis



Collecting, correcting, and normalizing performance data from a network's myriad of content distribution partners.

In television's "new normal," networks need quick answers about how shows are performing not only on their own domains but also among a range of distribution partners. That requires a data environment that facilitates information sharing across organizations, produces standard, accurate data, and can be accessed by frontline managers and senior executives alike.

Background

A large, well-known network has multiple brands, all of which create original content that is consumed on its own domain, as well as distributed to myriad Multiple Video Programming Distributor (MVPD) partners, such as Comcast and Verizon, as well as over-the-top (OTT) providers, such as Roku. In this "new normal" content distribution network, the content distribution team needed to answer key business questions, such as:

- How are individual shows and episodes performing on each partner's domain?
- How does that performance compare to the business terms agreed to by each partner?
- How does that performance compare to the performance of its own domain?

Additionally, the team wanted to plot show performance over time, as well as identify the ebbs and flows of consumption within set periods of time, such as business quarters.

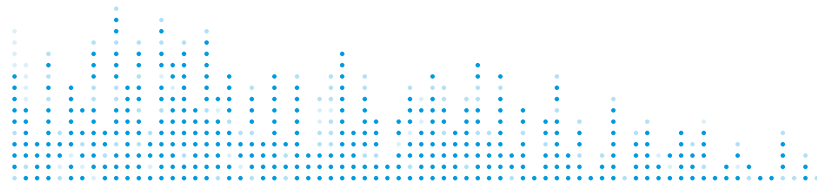
The network needed to pull viewership data from a multitude of MVPD and OTT providers, centralize it, and come up with a way to do comparisons that people with minimal data-science skills can understand.

Challenge

Data from the network's existing system often contained variations and errors, so they needed to collect data from its MVPDs and OTT partners directly. Many offered an API for data collection, but many didn't, which meant a web-scraping solution, customized to each partner, would be required.

Their data was also far from standard or accurate. All sorts of data fields, such as show titles, varied from partner to partner, and often contained manual errors. Complicating matters further, the network lacked a nomenclature for standardizing data. For instance, there was no authoritative definition – much less an authoritative ID – for shows and episodes.

There were other inherent challenges with the dashboard itself. To begin, the network wanted one tool for both frontline managers and senior management. The managers wanted a visualization tool that would allow them to answer a range of specific questions about their shows and episodes across partners, while senior management wanted to drill down into the data based on a different set of criteria. That meant creating a single dashboard that allowed both constituencies to get the answers they needed quickly and easily.



Solution

We implemented a multi-step solution to achieve the client's goals.

Step 1. Data Retrieval from Partners. Because the client was interested in show- and episode-level data, we needed to get raw data from each partner. In some cases, the partner offered an API which we could use to collect the data. In the cases where the partner lacked an API, our data engineers wrote custom web-scraping scripts to collect the data. Since there were dozens of scripts to write, we opted to use Python, which offers a rich library ecosystem and is quick to write and is very scalable and readable. In fact, we could write Python code in half the time it takes to write code in Java, which saved the client time and money.

Step 2. Centralize Data in a Big Data Environment. Once the data was retrieved from each partner, we imported it into a Hadoop Hive environment within the network's private cloud. This environment was robust enough to handle the tasks of normalizing and analyzing the data, as well as offered the security the network wanted.

Step 3. Creation of a Standard Nomenclature. Because content distribution via third parties is the new normal of networks, we were keen to create a standard nomenclature for show and performance data, regardless of whether it stemmed from the network's domain or from a partner's.

Step 4. Correct and Normalize All Data. Once the nomenclature was created, we normalized all of the raw data from all of the partners against it. For instance, we eliminated variations in show names (e.g. Show A vs. ShowA), as well as mis-categorizations, such as determining that the record is really attached to the correct show. This also applied to episode names. This was a massive data quality exercise, requiring us to eliminate inconsistent formats between thousands of datasets. Upon completion, we exported the cleaned up data to the network's enterprise database so many departments could use it.

Step 5. Create Interactive Dashboards for Analysis. Finally, we created persona-based interactive dashboards

using Tableau Server, a solution that allows over 400 users to access and interact with the reports and download copies to their own laptops via a browser. Put another way, Tableau Server made these reports available at scale, even though none of the business users had Tableau software on their laptops.

The dashboard also included a deep palette of filtering so that each user could easily get the insights of most interest to them. The content-distribution and research teams could easily query the data for business answers, such as:

- How is my show trending period over period?
- Regardless of scale, are we up or down quarter over quarter?
- What is the net impact on my viewership?

For senior managers, data was aggregated up to higher levels.

How Did it Turn Out?

Although retrieving and munging all of the data was a huge and time-consuming process, the network was able to see the first of its analysis within six months of engaging us. For the first time ever, they now have a complete picture of their digital/VOD business. This has enabled the content distribution team to assess which partners are top performers by show, as well as track if contractual obligations are met.

And thanks to the interactive dashboards, the research team is also able to understand the ebbs and flows of their viewership, and to get a complete picture to report to the sales and editorial team as needed.

The client was so impressed with the initial dashboard and the insights it revealed that they retained us to continue the project.

Contact

Nick Amabile | CEO | DAS42

nick@das42.com | 844-333-4232 | www.das42.com

