



\$2.5 MN saved by migrating from on-premise to Google Cloud Platform

An advertising exchange was looking to optimize operational costs on their existing on-premise infrastructure. The approach to streamline complex ETL processes and migrate to Google Cloud Platform improved operational efficiency and led to significant cost savings.

Business Challenge

The client operates the world's largest advertising exchange, connecting thousands of brands with consumers globally. They wanted to improve profitability by cutting down on operational costs involved in maintaining the existing on-premise infrastructure. With the rising volume of traffic across websites and apps, their infrastructure had evolved into complex interconnected systems, making it difficult to scale with different sources and ETL pipelines. This resulted in longer query processing times, with some ETL processes lasting up to eight hours, as well as inefficient manual data analysis processes. They wanted a scalable cloud infrastructure to:

- Enable faster data processing
- Improve data accuracy and proactively predict ETL bottlenecks
- Provide high resistance to failure
- Optimize maintenance cost

Client:

A large independent AdTech provider, USA

Industry:

Advertising Technology

Focus Area:

Cloud Migration,
Data Warehouse

Overview:

An advertising exchange was looking to reduce operational costs on their existing on-premise infrastructure. Sigmoid streamlined and processed 50 TB+ daily auction data and implemented the migration to Google Cloud Platform, resulting in the following benefits:

- Minimum manual intervention
- Better fault tolerance
- Low maintenance costs

Sigmoid's Solution

Sigmoid evaluated the existing processes to better understand inputs, business logic and outputs that formed the entire client ecosystem. A new ETL framework was developed and existing databases were streamlined, sending data from ad exchanges to a centralized repository warehouse - Google BigQuery. Sigmoid developed an Apache Spark-based ETL framework, hosted on Google Cloud, to set up new pipelines within minutes and manage feeds running in real-time. The code was migrated from Python to Scala, and Apache Spark was leveraged for data processing efficiency. By using open-source tools, libraries and documentation, the existing pricing structure was simplified to charge only on actual usage. Google's Cloud Dataproc, a fully-managed cloud service for running Apache Spark was also used to allow preemptible virtual machines that lowered the compute costs for non-critical data processing.



Business Outcomes

\$2.5MN

Annual cost savings

15x

Faster data availability

2.4x

Improved data accuracy

150BN+

Rows of data ingested daily from over 120 feeds

The migration, which was completed in under 20 weeks, ensured that the client system required little manual intervention and was easy to debug, extend, and modularize, resulting in lower maintenance and better fault tolerance. Sigmoid also improved the data accuracy by nearly 2.4 times and ingested over 150BN rows of data daily from 120+ feeds. The overall solution resulted in \$2.5 MN annual cost savings and improved customer satisfaction with 15x faster data processing.

Let's Get Started!

If you're looking to solve data challenges in your on-premise or cloud environments, contact Sigmoid's experts at marketing@sigmoid.com.



About Sigmoid

Sigmoid is a leading data solutions company backed by Sequoia Capital. We offer best-in-class services in the end-to-end data value chain spanning across Data Science, Data Engineering and DataOps. Our consultative data-first approach has a proven record of accelerating the digital transformation journeys for global F500 enterprises. With expertise in open-source and cloud technologies, we specialize in building solutions at speed and scale that consistently translate to successful business outcomes.