Company-wide operational analytics

Appsflyer is a leading mobile marketing analytics and attribution platform, the source of truth for many leading brands when it comes to attribution. Over 12,000 companies rely on Appsflyer to help them track, analyze, and understand buyer journeys, from tracking mobile installs to ad interactions across mobile and other channels. Appsflyer works with over 6,000 partners to make tracking buyer journeys a reality.

The challenge - add speed, scale and agility for internal analytics

Over the last few years, Appsflyer had pushed its internal analytics to the limit.

Three years ago, Appsflyer started using Looker. Over time, adoption grew to 1,000 Looker users across sales, support, marketing, finance, and HR, including 60 Looker experts who supported other users by building and maintaining the various dashboards.

Appsflyer was running their analytics on Amazon Athena, but as usage had grown, so had the challenges. First, the data had become too big for Athena to handle. Appsflyer has 35 petabytes of terabytes of raw data about customers and their activities, with multiple tables consisting of billions to hundreds of billions of rows. Athena could not handle more than 5 billion rows of data in any query. So they had to aggregate the raw data using Databricks to a smaller data set that Athena could handle. Athena also could not handle more than 20 concurrent queries, which had become a problem by the time Appsflyer reached 1,000 Looker users.

Second, the ETL process was slow, expensive, and inflexible to change. A spark job often took weeks to change, and days to run if they had to rebuild all the dimensions and history. This was not only too long, since users often required new reports in days, but also very expensive. Updates also became a problem with GDPR regulation. Each time an entity opted out, Appsflyer had to rerun entire Spark jobs to remove them, which not only took a long time to run but was very expensive as well.

Third, even with the dramatically reduced data sets, Athena was still too slow. Some queries took minutes to run. Most interactive analytics, such as the executive dashboards, required query times of a few seconds at most.

When the company needed a new report that required the detailed results from a join of two multi-billion row tables, Appsflyer decided it was time to replace Athena.

The solution - Firebolt cloud data warehouse
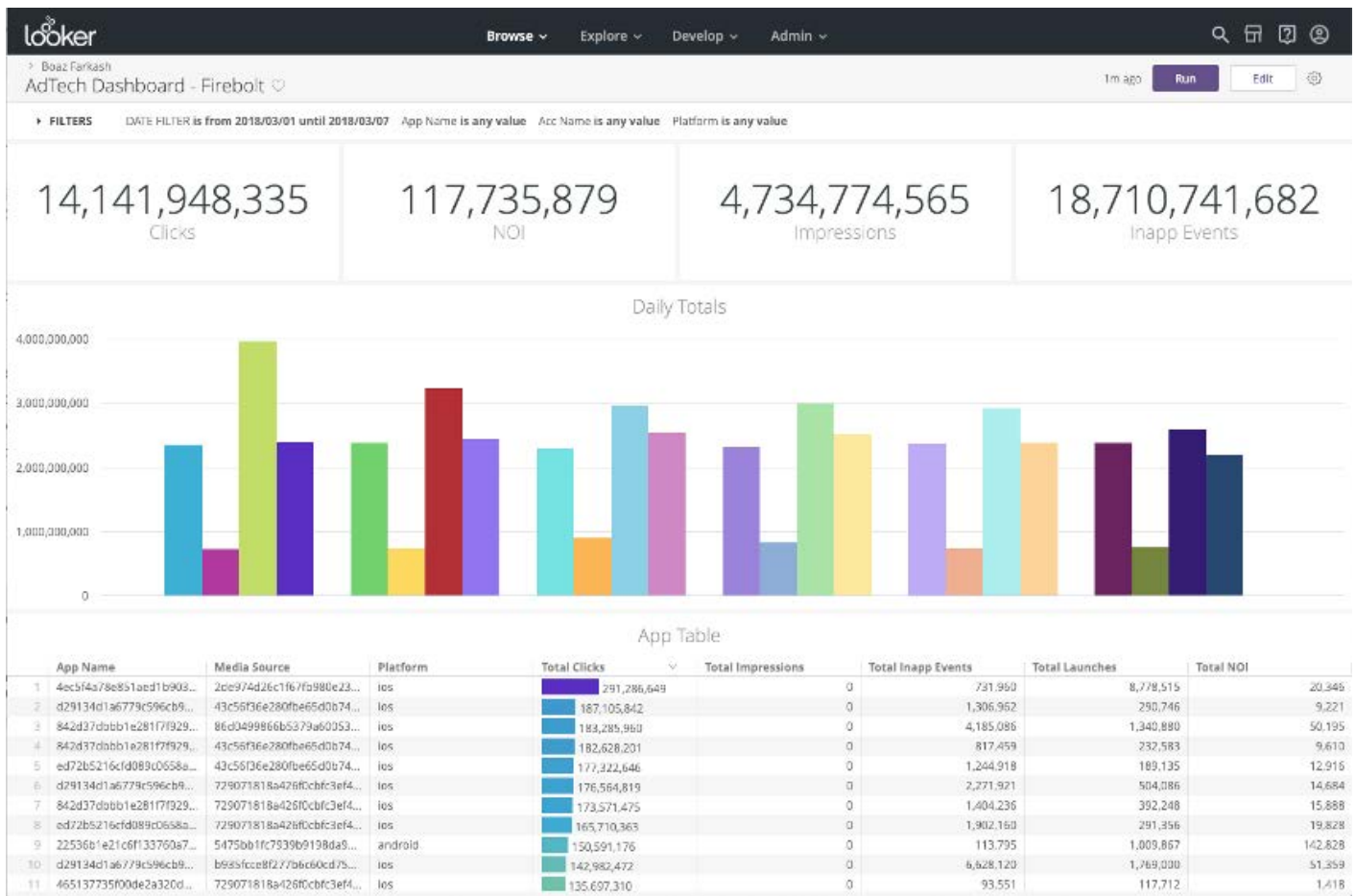
Appsflyer started to look at a cloud data warehouse that could store all of their data and aggregations, deliver the performance they needed, and provide greater self service so that various teams could deliver new analytics faster.

Appsflyer evaluated several vendors in a proof of concept (POC) with 3 stages: joining a big and small table, joining two big, multi-billion row tables, and a dashboard use case. They were already using BigQuery, and evaluated Snowflake, Vertica, Dremio, and SQream. They had evaluated Redshift a while ago and had ruled it out. They also used Druid with a customer Web interface for their customer-facing analytics and decided against using it because they needed a more general-purpose data warehouse.

In the end, they chose Firebolt based on its much faster performance, and much better price-performance. With Firebolt, all of Appsflyer's 1,000 Looker users are now able to run any analytics, at any scale, in seconds, and deliver new reports within days.

Implementing the initial project

As part of evaluating Firebolt, Appsflyer took the same schema and queries from Athena syntax and ran them on Firebolt, performing joins with 3 billion and 26 billion row tables.

Before Firebolt, any new data or schema needed to be processed using Spark. It could take weeks to develop and run the Spark jobs. With Firebolt, data engineers can now do most of the ELT in hours using 100% SQL, and run it in Firebolt at any scale. The first project took 2 weeks or so of development. Most work now takes a few days or less.

Run queries at any scale in seconds or less

Performance was 10x or greater from the start, even without tuning. During the POC, the smaller join on Firebolt, which had been taking at least 40-60 seconds using Athena, was under 5 seconds. Even the larger join, which could not be done on Athena, ran within 5 seconds after a little optimization. Now even the larger joins can be run in 1-2 seconds, and filtering is millisecond

	Athena (sec)	Firebolt (sec)	Performance boost
Single left join - last 3 months	120	1.8	x67
Filter by time - last year	36	1.1	x33
Filter by time - last 3 months	11	0.3	x37
Filter by media source	104	1.4	x74
Filter by App_id and media	26	0.5	x52
Filter just by App_id	73	0.4	x183
Count distinct on all data	117	2.1	x56

Support any data, query, and user scale

When Appsflyer was using Athena, they could not analyze any data beyond 5 billion rows, which prevented them from performing a broad range of analyses. With Firebolt, there is no limit. Appsflyer is now able to analyze any level of detail, up to hundreds of billions of rows, with confidence. Firebolt's ability to scale horizontally also meant it easily supported any number of concurrent users.

Up to 10x lower cost

With Athena, cost had not been as much of a concern; it just didn't work at scale. But it was a major concern as they were evaluating their alternatives, in part because of their scale. BigQuery ended up

being the other option they considered. Beyond Firebolt's faster performance, it also ended up being up to 10x lower cost, in part due to its greater efficiency per node.

The future - an even more data-driven Appsflyer

Before Firebolt, analytics at Appsflyer had been throttled by Athena's limitations. With Firebolt, those limits are gone. With Firebolt, employees can now have interactive conversations with their data. They can use Looker to analyze data as they are trying to make decisions because everything in Looker now takes seconds or less.

As a result, customer service and other groups have already started to expand their use of Firebolt and Looker. This includes analyzing much larger data sets, including more granular data, and longer time periods.

"

With Firebolt, our 1000 Looker users can now run analytics against billions of rows and terabytes of data, in seconds or less.

Alexandra Sudilovski Senior BI Expert & Looker Guild Master, AppsFlyer