Industry: Media

Publicis Media: Moving Forward in a Big Data World

Business Challenge

In 2016 multiple teams from the various media agencies merged to form Publicis Media. This merger revealed the need for a central data and analytics platform. "We wanted our agency teams to be able to mine data, but not to have to deal with the operational overhead of managing data infrastructure," explains Darren Smith, who leads the engineering and data teams. "Our intent was to democratize data."

According to Smith, the team's existing data infrastructure "was a bunch of bespoke solutions" that combined AWS Redshift, large monolithic on-premise servers, and various unwieldy traditional technologies. Offering a central data and analytics platform would require both a complete overhaul of this infrastructure and some way to tie all of its pieces together.
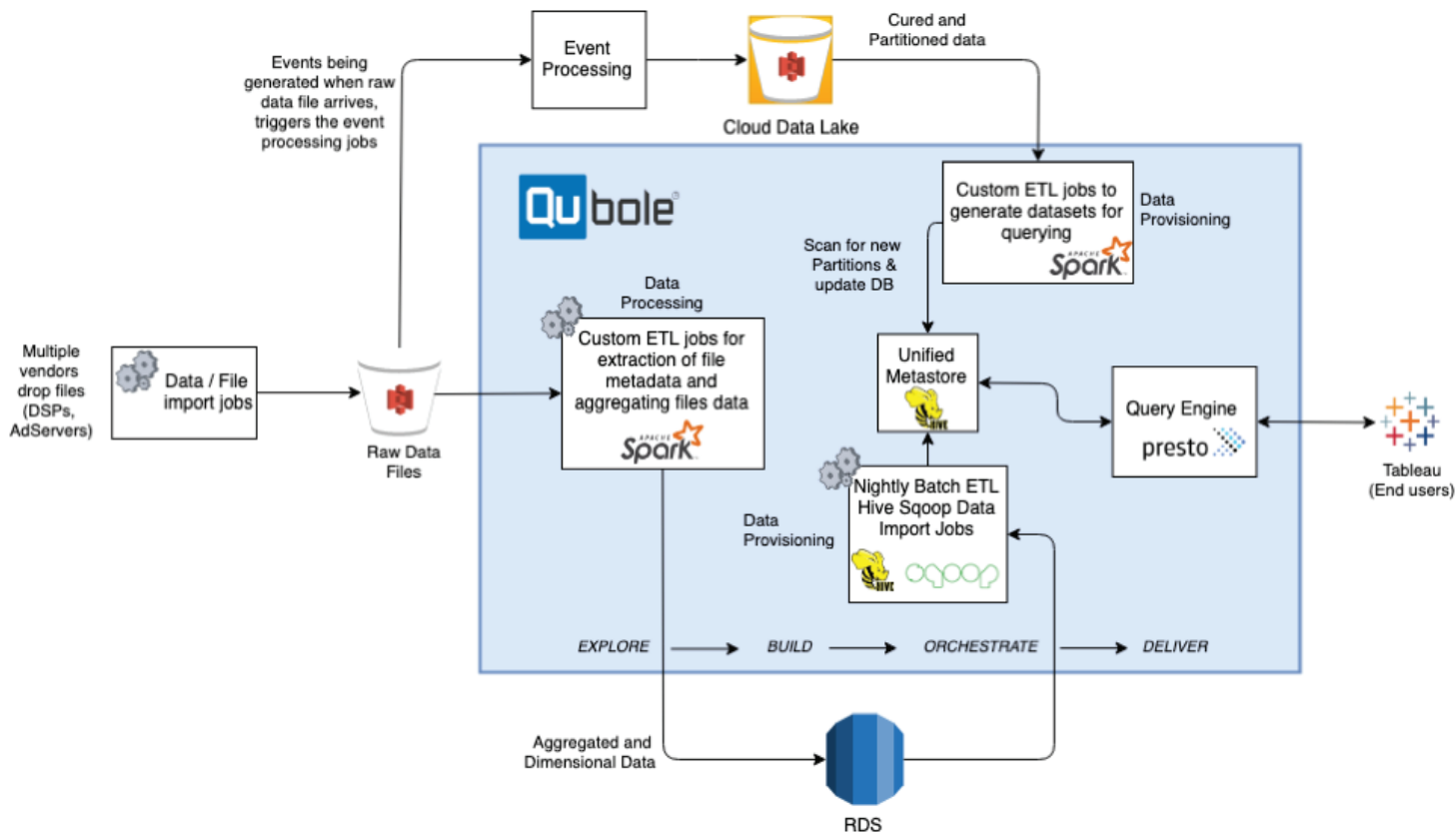
About Publicis Media

Publicis Groupe is one of the four solutions hubs of Publicis Groupe [Euronext Paris FR0000130577, CAC 40], alongside Publicis Communications, Publicis.Sapient and Publicis Health. Led by Steve King, CEO, Publicis Media and COO, Publicis Groupe, is comprised of Starcom, Zenith, Digitas, Spark Foundry and Performics, powered by digital-first, data-driven global practices that together deliver client value and business transformation. Publicis Media is committed to helping its clients navigate the modern media landscape and is present in more than 100 countries with over 23,500 employees worldwide.

> " *The focus of our team was to build a data architecture and infrastructure that would allow our agency teams to move forward in a big data world."*
>
> **Joe Tan**
> Director of Products, Publicis Media

Events being generated when raw data file arrives, triggers the event processing jobs

Event Processing

Cured and Partitioned data

Cloud Data Lake

Multiple vendors drop files (DSPs, AdServers)

Data / File import jobs

Raw Data Files

Data Processing

Custom ETL jobs for extraction of file metadata and aggregating files data

Spark

Custom ETL jobs to generate datasets for querying

Data Provisioning

Spark

Scan for new Partitions & update DB

Unified Metastore

Query Engine

presto

Tableau (End users)

Data Provisioning

Nightly Batch ETL Hive Sqoop Data Import Jobs

EXPLORE → BUILD → ORCHESTRATE → DELIVER

Aggregated and Dimensional Data

RDS

A Centralized Platform for Democratizing Data

"The focus of our team was to build a data architecture and infrastructure that would allow our agency teams to move forward in a big data world," says Joe Tan, director of products at Publicis Media. The resulting infrastructure couples a global data lake —which stores large volumes of multiple types of data—with a framework to ingest and process data. It consists of a range of technologies on Amazon Web Services (AWS).

In addition to building this data infrastructure, Tan's team had another job: "to provide tools that allow agency teams to really focus on doing analytics for their clients instead of having to worry about data ops and data engineering." For this, the team turned to Qubole. Qubole enables agency teams to "work with the data they're used to in the tools and languages they're used to, like Tableau and Presto, or SQL, Python, R, Scala, etc." says Tan. It also helps Publicis Media make data available to users with different skill sets. "It even," says Tan, "gives users the ability to learn how to do more with minimal additional effort."

> *Qubole really meshed well with the overall architecture and design of our data lake. I don't think we could have found a better platform."*
>
> **Darren Smith**
> Engineering and Data Teams Leader, Publicis Media

As more and more clients have grasped the potential power of Publicis Media's platform, Qubole has played a key role in helping increase its adoption. "We have had a steady growth rate of one to two agency clients onboarding onto our platform per month," says Tan. "That might not sound like a lot, but a lot of those teams service multiple clients of their own, so it's pretty impactful."

Scalability Is Key in a Big Data World

Publicis Media handles lots of data for its agency clients. In fact, its data lake stores close to a petabyte of it. Agency clients use data in this data lake to run machine learning models for analytics purposes. Some data sets used in these models are massive; others, not so much.

Before Qubole, scaling to process larger data sets posed a challenge. "I regularly walked into offices and ran into someone who'd had a model running for six hours," recalls Tan. Qubole solves this problem by enabling agencies to automatically scale up compute infrastructure for large jobs, and to aggressively scale back down when a job is complete to keep costs low. So, jobs that once took six hours to complete can now be finished in mere minutes, with almost 10,000 queries per month on average. In addition, Qubole also supports multi-region data availability without latency—further improving the performance and consistency of Publicis Media's data globally.

This ability to quickly glean important insights from data has resulted in a shift in thinking among some agency clients. Before their media planning and buying efforts were largely reactive. Now they are becoming more prescriptive. Tan expects more and more agencies to adopt this new mindset as the advantages of analytics continue to crystallize.

> " *Our end-users are just excited to have a scalable solution that's centrally managed."*
>
> **Joe Tan**
> Director of Products, Publicis Media

Boosting Administrator-to-User Ratios

Qubole isn't just powerful. It's also remarkably easy to administer. "Running a small team of just one administrator and two solutions architects," says Tan, "I can support in the range of a hundred users." Using any other similar tools on the market, "I don't think I'd be able to make that statement," Tan adds.

Of course, data security is top of mind for Publicis Media. Qubole addresses its requirements with regard to single sign-on, strict role-based access control, and agency data isolation, among other security issues. While both Smith and Tan see these features as "table stakes," Smith acknowledges that, "A lot of vendors don't support them."

Looking Ahead

For Publicis Media, using Qubole has already produced remarkable results. But so far the group has used only a few of Qubole's features. "Looking ahead, our priority is to investigate some of the features on Qubole that we aren't using now," says Tan. Specifically, Tan expects to start exploring use cases for Apache Airflow on Qubole for workflow management purposes.

Business Value

• A central data and analytics platform that democratizes data

• Ability to manage nearly 1 petabyte of data

• Reduction in model run time from six hours to mere minutes, with almost 10,000 queries per month on average

• Multi-region data availability without latency

• Easy administration

• An administrator-to-user ratio of 3:100

• Support for robust security and compliance requirements

For more information:

Contact:
sales@qubole.com

Try QDS for Free:
https://www.qubole.com/products/pricing/

469 El Camino Real, Suite 205
Santa Clara, CA 95050
(855) 423-6674 | info@qubole.com

WWW.QUBOLE.COM