Meeting the Challenge of Big Log Data Analysis for Marketers

AuriQ Essentia Platform
Whitepaper

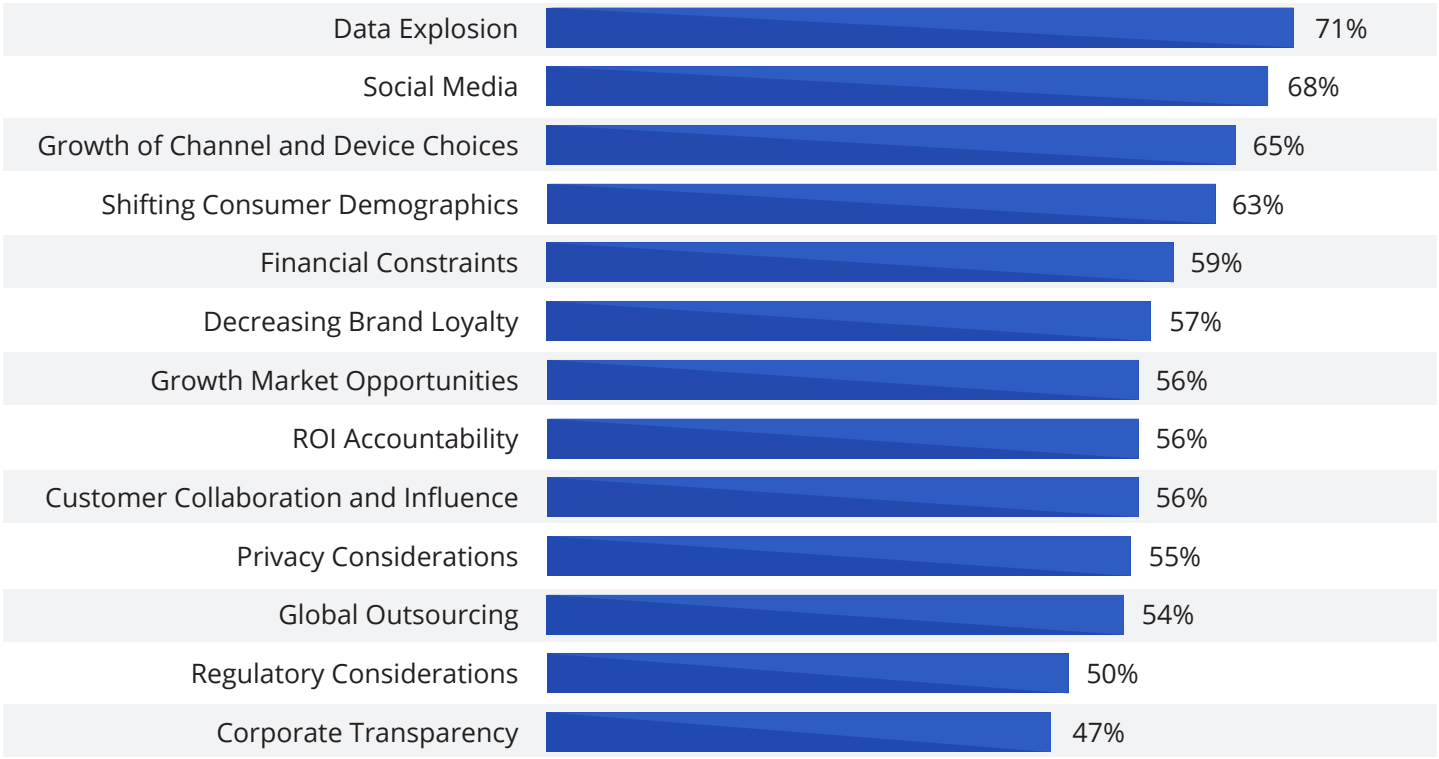
AuriQ Systems Inc.
Pasadena CA

Data Explosion: CMO's Biggest Pain

The modern CMO is faced with a rapidly changing marketing and advertising landscape. There are more channels, more paths, more methods and an ever-growing audience, whose behavior is always changing. Through all of this marketers are trying to gain deeper understanding into their customers' journey. Advances in technology have made it possible to track almost everything, which has the potential to give unprecedented visibility into this journey. The challenge is overcoming the uncertainty in dealing with the vast amount of data generated.

According to a survey in mid-2013, "Data explosion" was listed as the biggest pain of CMOs. 71% of 1700 enterprises named it as the top concern.

Percent of CMOs Reporting Under preparedness



Source: IBM whitepaper June 2013

But, why is big data such a problem? Simply put, it's expensive to analyze. Marketing has become a data-driven activity, improving success, understanding customers, validating spend, each of these rely on, and are augmented by, the availability of data. Today's marketer has a tremendous amount of data available, but are finding that processing such huge amounts of data in a timely manner not cost effective using prevailing technologies.

Barriers to Using Technology

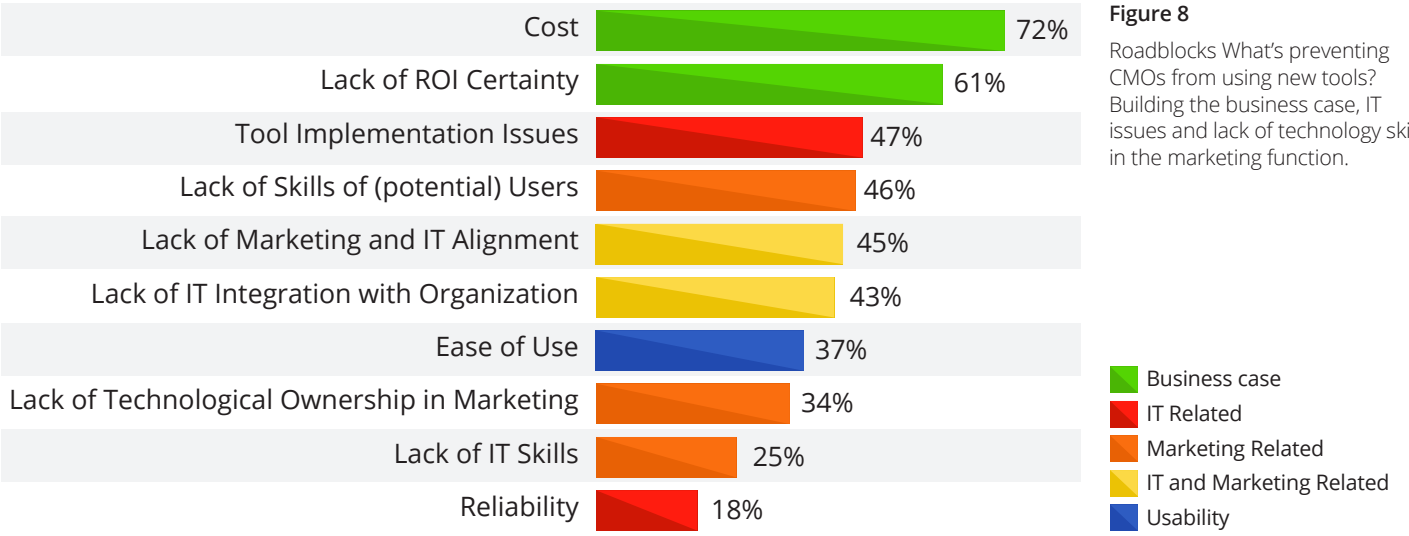


Figure 8

Roadblocks What's preventing CMOs from using new tools? Building the business case, IT issues and lack of technology skills in the marketing function.

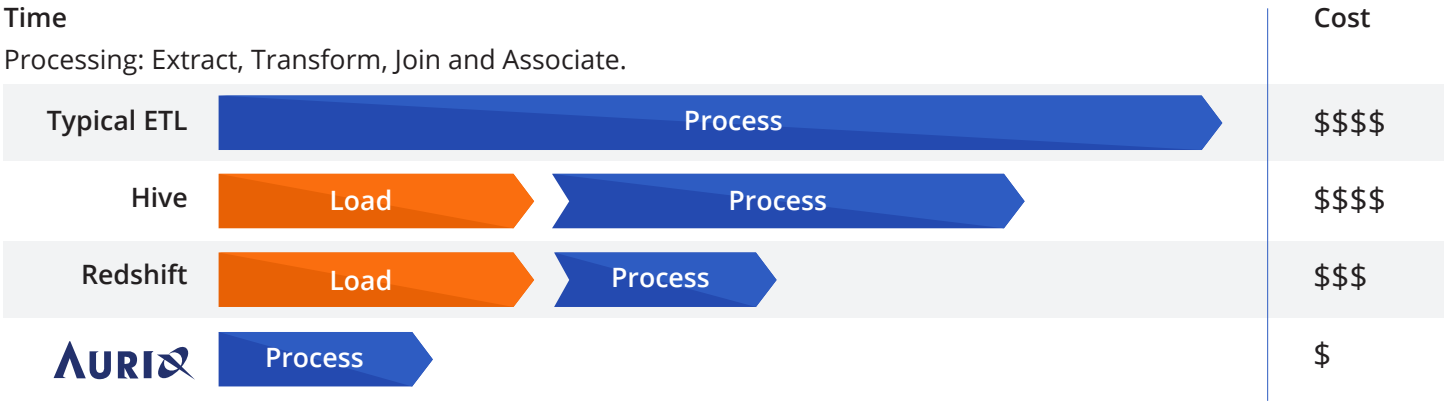
Current Options for Data Processing Fall Short

There are already a few options out there for processing and analyzing this type of data. Hadoop, for one, is a great general-purpose platform for analyzing data. However, using Hadoop to process big log data is a little like carving up a T-bone steak with a Swiss army knife. Sure, eventually it will work but not nearly as well as a good steak knife.

Current options like Hadoop, Redshift, and BigQuery, rely heavily on writing data to disk, this means that the processing and analysis speeds are slow and resource heavy. That's fine when you have a small data set and time and cost are not priorities. AuriQ Essentia, by contrast, is designed specifically for speed and scalability. By emphasizing the use of memory instead of writing to disk, the speed of processing, data cleansing, and analysis are all dramatically increased.

As big log data becomes a crucial factor in marketing intelligence it's important to utilize a tool that is specifically designed for the task.

Time & Cost to Process Terabyte Size Log Files



AuriQ Essentia (Big Log Processing Platform)

AuriQ Essentia is a brand-new, specialized data processing engine designed for fast, cost-effective analysis of big log data. This platform allows for analyzes of big log data over 10 times faster than current standards, consistently outperforming the fastest existing solutions, such as Redshift. And, it can analyze big log data more than 100 times faster than the typical Hadoop based systems.

Modern marketers increasingly need to attribute their successes and failures to specific data gathered from a variety of sources. The growing need for advanced attribution modeling requires high-speed analysis. However, the high cost and lengthy processing times of most current solutions often precludes sufficient analysis.

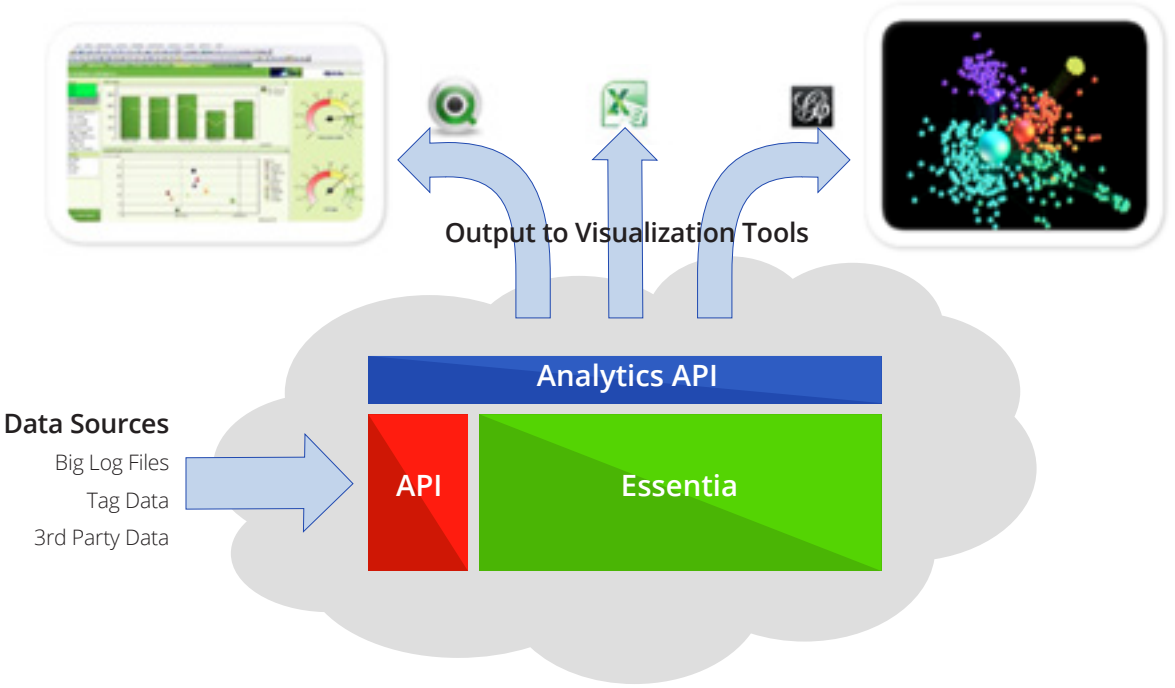
AuriQ Essentia enables advanced marketers to perform advanced attribution modeling by cutting the cost and time to process big log data significantly.

This is a cloud-based service on Amazon AWS, so any marketer can use it without heavy initial investment. Various templates are available that allow non-technical users to do advanced attribution modeling without any programing skill. It's focus is to reduce big log data into a structured data set so marketer can use simple tools like Excel, Qlikview and any other BI tool for further reporting

Attribution Analysis

Understanding the customer journey is a nuanced and complex process. A typical consumer will be exposed to a number of ads prior to making a purchase, and understanding this connection is crucial to correctly attributing the success of marketing efforts. However, the amount of data produced by online marketing is staggering. Ad server logs often run into billions of rows of data on impressions, clicks and more. Attempting to combine those with web server logs (also often in the billions of rows themselves) is challenging to say the least. Because of the time it takes to properly correlate and analyze this data it is difficult to extract actionable intelligence.

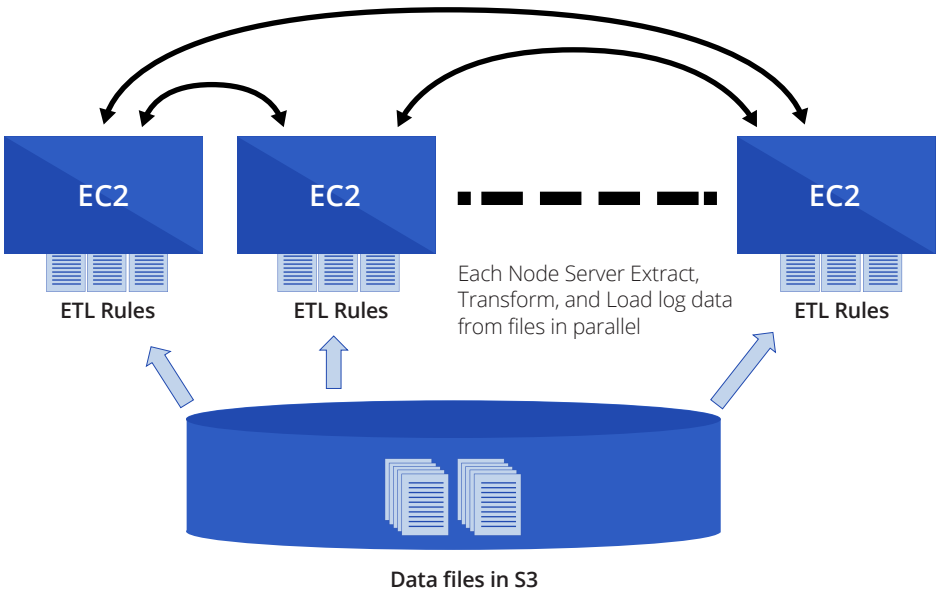
AuriQ Essentia solves this problem by considerably reducing the amount of time required to analyze the data. And, our experience with this process has yielded a unique ability to "match" users from one data log to another, thus creating a complete picture of the customer journey. What used to take weeks, now takes days. What took days, now takes hours, and the result is actionable intelligence in time to improve active campaigns.



Parallel ETL

AuriQ Essentia utilizes a powerful ETL to extract, transform and load useful info from unstructured data. It can run on a cluster of servers to easily scale up the processing power and speed.

Each Node Server communicates to each other to pass data around to set the right data in the right server.



Parallel User Profiling

User profiling is an intermediate data processing step for advanced marketing analysis. Examples of profiling can be user based such as “flag a user who used a specific search keyword AND looked at particular banner ad AND then purchased” and event based such as “flag events only between first organic search and first time to reach a particular page on the web site”. The challenge is that advanced marketing analysis requires creating a large number of profiles to analyze right data precisely. AuriQ Essentia makes such complex profiling very easy and very fast.

Parallel Query Processing

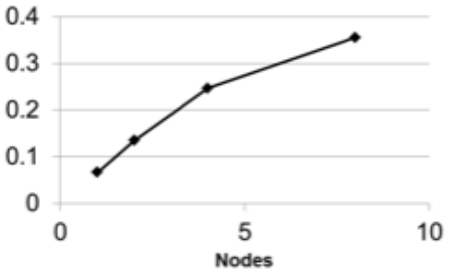
Running queries is also simple. A basic command line interface allows users to request AuriQ Essentia to filter, segment and aggregate data to produce a data set for reporting. Data sets are created in csv format so that users can import them directly into Excel or other reporting tools.

Scalable Performance

This shows scalability of Essentia. For both ETL and Query, processing performance increases almost linearly as the number of nodes increase. A small sign of saturation is seen on ETL at 8 nodes due to I/O speed of one bucket of S3. When the original data was distributed, there is virtually no saturation on the performance.

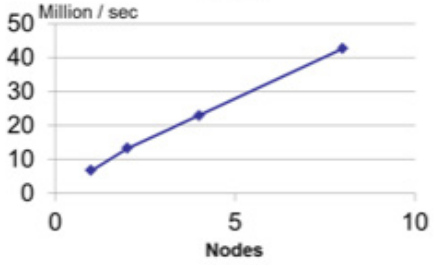
ETL

Extract, Transform log data from compressed files and Load them into BLP



Query

Simple filtering and aggregating



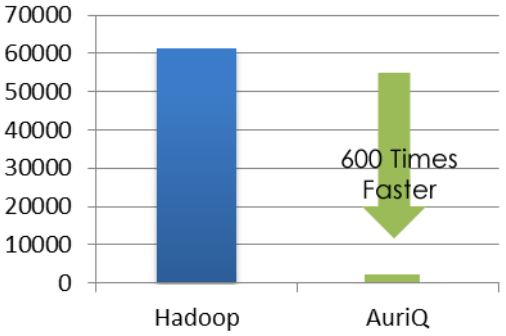
Benchmark

Benchmark results on real data.

A case to analyze 40GB zipped CLD

ETL

A Typical Query



BLP is much faster than Redshift. Saves time and cost!

When compared against Hadoop and Redshift in the areas of query processing and ETL time, AuriQ Essentia easily out performed the competition. As demonstrated in the chart, query processing was more than 100 times faster than Hadoop and ETL preprocessing and loading took significantly less time than with Redshift.

ETL

Unlike Redshift, AuriQ Essentia does not require preprocessing. It cleans the data and transforms as it reads the data directly from the file.

Query processing

Redshift is a great general-purpose super database and can process simple queries very fast, even faster than AuriQ Essentia. However, for complex queries that require multi-level JOIN, the performance is not great. For the benchmark test, we used typical queries, mid-level complexity, and found that AuriQ Essentia perform 5-10 times faster than Redshift. AuriQ Essentia's performance improves as the complexity of the query increases.

Interested in learning more about AuriQ Essentia big log processing capabilities, or do you have a project in mind already?

Contact us for a free consultation.

We'll assess your data analysis needs and help you determine whether or not AuriQ Essentia is the right option for you.

AuriQ Systems, Inc.
301 North Lave Ave, Suite 1001
Pasadena, California 91101
(626) 564-2781
essentia@auriq.com