

# CLOUD-BASED DATA ANALYTICS

## BUSINESS OBJECTIVES

Client is a cloud-based data analytics company focusing on operations, security and BI use-cases. It provides log management and analytics services that leverage machine-generated big data to deliver real-time IT insights

## SOLUTION

Our client has multiple sources from which marketing data is captured. Organizing the same data set from all sources into one space was challenging. The immediate challenges included:

Extracting marketing data from one of the sources via API. There is no concept of Pagination to collect all the data in one run. All the data will be collected internally, which goes through several internal loops until it won't expose all the records. This was complex to achieve in ETL tools and reduced the overall performance. Also, for one of the subjects, multiple iterations were required.

All these made the processing time taking and infeasible. End-to-end automation, use of variables, and multiple looping made the standard solution complex. So, error handling was also complicated.

## SUCCESS CRITERIA & BUSINESS VALUE

The standard solution was using a POST request in Matillion to get the 'Access Token' which was used as an endpoint to send the GET request to extract data. As mentioned above, there was no concept of pagination, so the endpoint was required to be looped through using variables to extract all the records. These made the standard approach infeasible. We used the Python script component in Matillion, which included the Python code having all the loops and variables, making the other mapping part simple. For failure recovery of data, we overlapped the data extraction time period. We used Change Data Capture(CDC) in ETL to filter out the duplicate records which came because of time period overlap. This solution produced the best results, reducing the time and complexity to negligible.

## CONTACT US

Infometry INC

+1 (510) 770 6400

+1 (510) 793 5859

39111 Paseo Padre Pkwy, Suite 203E,  
Fremont, CA 94538, USA

## TECHNOLOGIES USED

Matillion, Change Data Capture