

Smart customer segmentation case study.

We move through our digital worlds and leave crumbs of our digital DNA all over the place: A photo or a like in Facebook, a Tweet message, searches in Google, purchases in Amazon and other retailers, browsing histories. We also have many interactions inside the mobile applications and web pages of our many vendors and providers.

The smartest companies of our digital age are racing to compile all these pieces of info with the main objective of creating the most accurate and relevant profiles of their customers. A detailed customer profile, combined with the many ubiquitous predictive analytic technologies allow these companies to know how we like to be approached, what we think, what we like, and in some cases what we will do in the near future.

Here are some insights on how we had helped one of our telecomm customers to use Amazon AWS cloud technologies to create a system for smart customer segmentation on millions of users through many billions of data points. Amazon cloud gives us huge computer processing power on demand, at very reasonable prices.

We collect data from as many sources as possible. Some of the data is public: Twitter (we use Kinesis to listen to the hashtags we care about), in other cases we pull data through APIs (Clickstream, Google analytics). The rest of the data comes from the many interactions that the company has with the users through web sites and mobile apps as well as internal billing, POS systems and call centers.

The data is all collected and stored in S3, several on demand Spark clusters are instantiated to ETL, clean and merge the information. The resulting cleaned and tagged data is written back to S3 and some data pieces are used to feed a “live” data warehouse in Redshift, that is available 24/7 for analytics and reporting.

A different Spark process for segmentation run at set intervals and uses the clean data to produce anonymized result sets in which each customer is assigned a unique identifier and classified in as many segments as we have defined, we segment based on gender, age, product usage, expenditure, number of contacts, behavior, and many other variables. The result of the segmentation process is persisted and tagged with the date in which the segmentation ran.

The persisted segmentation table allows for immediate access to anonymized, relevant and current customer segment information that is directly interfaced with campaign management systems and also used to feed predictive analytic models.

The resulted segmented data is stored into a dynamoDB table and we created a server less web service that uses Lambda to retrieve the segments information for a specific customer in fractions of seconds. This web service is used by CRM systems to obtain a 360 view of the customer data.