colt

Client Name	Sector	Challenge
DefinedCrowd	Big data	To provide a safe and reliable high-speed connection to access a large volume of data
Business	**Products and services**	
Training data supplier for artificial intelligence	IP Access	

The start-up DefinedCrowd has a platform that allows data scientists to collect, structure and enrich high-quality data to train artificial intelligence algorithms. The quality of the data handled and the speed of processing are key to its business.

Colt provides the company's connectivity so they can access all their data quickly and efficiently. Ricardo Gusmão, IT & Information Security Manager at DefinedCrowd, highlights Colt's international presence as a decisive advantage, and he sees Colt not only as a supplier but also as a partner to help the start-up in its growth phase.

DefinedCrowd enlists Colt to connect with its smart data platform

Everyone is talking about the possibilities of artificial intelligence, but most pay less attention to the driving force behind this technology: data. Specifically, data quality, which is essential to training AI algorithms. DefinedCrowd saw a business opportunity here and the company was founded in 2015 with the goal of creating an efficient system for turning unstructured training data into structured, high-quality datasets.
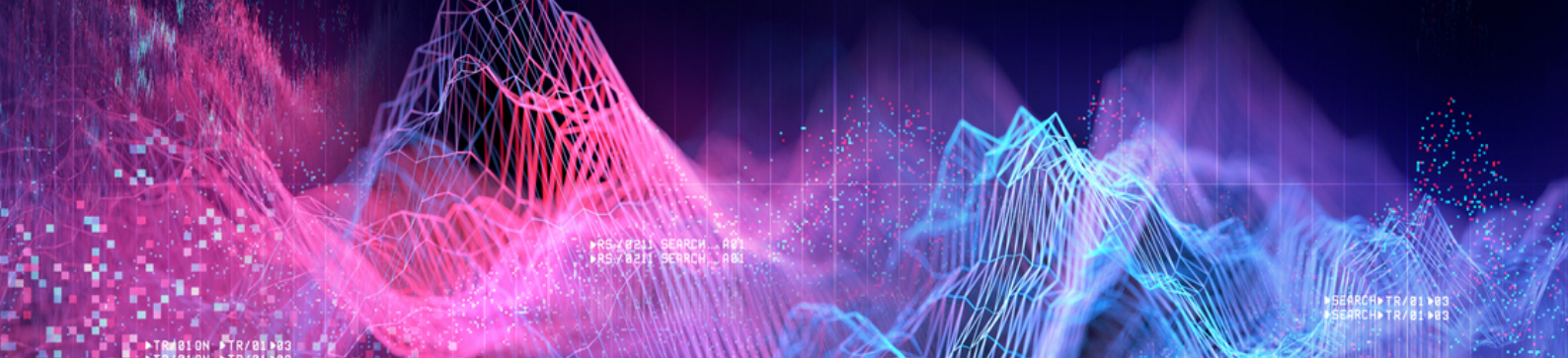
DefinedCrowd works with large volumes of cloud data that require reliable high-speed connectivity, which they entrusted to Colt. This service enables it to access and manage its smart data platform, which collects, structures and enriches data in order to train machine learning models. The source of the data can have two different origins. Clients can provide the data for DefinedCrowd to structure, or they can request the collection of data.

It is estimated that data scientists spend 80% of their time structuring data. The goal is to release them from that onerous task so that they can focus on creating high-performing AI algorithms. Catarina Salteiro, Global PR & Communication Manager at DefinedCrowd, explains how the system works:

> "Clients go to the platform and can customise the type of data they want. Let's say they are working on a voice recognition system and they need training data. They decide how many scripts they want people to read, how many times they want them read and in what languages. Afterwards, these requirements are transformed into microtasks which are funnelled to our human-in-the-loop community."

DefinedCrowd's community is a key part of the company's offering, with more than 130,000 users covering over 70 languages and dialects. Users tag images, read and annotate texts, as well as performing other specific microtasks that contribute to client projects. This, along with the expertise of its team of engineers, enables the company to excel in voice and natural language processing technologies. To complement their services, they also work with other machine learning ecosystems such as computer vision. "Our community does the microtasks, which enables us to get high-quality data sets. At this point, we have highly qualified members to provide the data our clients need, which allow us to guarantee 95 to 98% of quality," says Salteiro.

To access this entire data infrastructure, DefinedCrowd uses IP Access from Colt with

symmetrical 1 Gbps. With this connection, it manages its smart data platform and the entire volume of data with which it works. For productivity to be optimal, the data must be accessed at high speeds, yet other factors are also decisive.

Ricardo Gusmão, IT & Information Security Manager, stresses the importance of Colt's web security service which protects their system against possible attacks, as well as the SLA conditions and the reliability of the connection. However, there is one aspect which Gusmão particularly values: "Availability is very important, especially in companies like ours, who depend on communication. This availability, as well as reliability, are the most important factors, and Colt is capable of providing them". He adds: "I really see Colt as a partner, not just as a supplier".

DefinedCrowd has a long exciting road ahead of it, and Gusmão is convinced that Colt can be of service throughout its growth process. The transversality of its smart data platform allows it to work with clients in different industries, from customer care in the healthcare sector to automotive, energy, fintech, retail and the media. Its product strives to help companies improve the quality and scalability of their machine learning applications and accelerate their launch into the market.

Most of DefinedCrowd's clients are ranked on the Fortune 500, with prominent names including BMW, MasterCard, and Yahoo Japan, and it has solid ties with rising companies in artificial intelligence. Amazon is among their key investors, and recommend DefinedCrowd's services following their role in the development of Alexa. The company is a Microsoft co-sell partner and

> "Availability is very important, especially in companies like ours, who depend on communication. This availability, as well as reliability, are the most important factors, and Colt is capable of providing them". "I really see Colt as a partner, not just as a supplier
>
> **Ricardo Gusmão,**
> **IT & Information Security Manager**

their solution is also integrated in IBM's Watson platform.

DefinedCrowd's rapid growth has only accelerated further since its summer 2018 announcement of an 11.8 million US dollar investment round. It began 2019 with fewer than 80 employees and plans to end the year with 150. Right now, it has offices in Lisbon, Porto, Seattle and Tokyo – Colt has fibre optic communication in these cities and offers the possibility of having a single provider in all its global locations.

To support its future expansion, DefinedCrowd is considering the possibility of a dedicated cloud for Microsoft Azure, which they use for their services. "This is why having Colt present in the countries where we operate is so important, because the information access speed is extremely important to us", says Gusmão. "Being located in places where we have offices can benefit us when dealing with connections". These are the foundations for the goal the start-up has set for themselves: to be the number one training data supplier for artificial intelligence initiatives.

DefinedCrowd®