In 2001, Gartner defined big data as data that 1) contains greater variety 2) arriving in increasing volumes and 3) with ever-higher velocity. These are known as the three Vs.\(^1\) Almost two decades later, this definition still holds true. Data sets are becoming more complex, and contain massive volume of data which enable users to address business problems that they wouldn't have been able to address before.

In recent years, big data has become increasingly more important to companies. In fact, it has been said that 90% of all data has been created in the past two years.\(^2\) Why? The massive internet of things means that every phone, watch, car, desk, or kitchen appliance can collect and share data. The computer hardware needed to handle big data is being produced more effectively and affordably. At the same time, open source software from many developers has given way to a lower barrier to entry for companies looking to source data. This has led to a growing big data workforce comprised of individuals who have experience working on big data projects or have a degrees in these areas.

For Commercial Cards, data is particularly important. The data that flows through the Commercial Card infrastructure is powerful — it can be used to negotiate pricing with suppliers, show a company where their employees are spending, and to predict travel budgets for the upcoming fiscal year. We have all this data, but how do we harness its power and make it easily digestible to the end user?

Commercial Card transactions occur at tens of millions of merchants around the globe and historically there has been no existing process to correctly identify merchant names. This means that companies do not have full visibility into their overall spend with the merchant. Let's take Starbucks as an example. There are thousands of Starbucks around the world (seems like a coffee addiction is a commonality!), and when the merchant name is passed to the card issuer, the name can differ based on the particular store. A merchant name might include a store number, or vary from country to country, or indicate a different buying channel. So, while the cardholder might be buying the same latte at a Starbucks in Detroit or Brussels, the merchant name may come up slightly different on their statement. For companies, this can be challenging especially when trying to determine how much spend is going to the merchant or when negotiating rates with suppliers. In this case, the data is there but it’s not being utilized to its full potential.

---

1. [https://www.oracle.com/big-data/guide/what-is-big-data.html#link8](https://www.oracle.com/big-data/guide/what-is-big-data.html#link8)
That is where big data, artificial intelligence, and machine learning come in. In 2017, TTS established the AI Horizontal to deliver in-house data science solutions for business use cases. Commercial Cards partnered with the team to develop a proprietary merchant name cleansing algorithm that applies a rule based approach for certain merchants and leverages a machine learning clustering approach to improve every merchant name based on the transaction description. This means that companies have a 360-degree view of their spend by segment and industry, enabling them to better understand their spend and increase their purchasing power. In our example above, this means that the company is able to see thousands of "merchants" and transactions bucketed under one merchant name – Starbucks. For procurement departments, this means that they will be able to give definitive answers to questions such as:

- How much did we spend at suppliers like Starbucks, Staples, or Amazon last year?
- What are my top 50 suppliers that are paid by card?
- Who is no longer accepting my card payment and who has stared accepting my card payments?

There is no doubt that data will continue to become incredibly powerful for companies around the world. Emerging technologies will allow for new ways to leverage data, providing deeper insight and better conclusions. For Citi this means that we will continue to invest in refining our data so that our clients reap the benefit of all the data flowing through their cards program – ultimately creating greater insight into their overall business.