# Sharing Data SECURELY

**By Ankita Gupta**

Measurable metrics are a key element of goal setting and success. However, the days of simply using sales revenue or customer satisfaction ratings to determine success are gone. Today's technology and tools allow us to collect a lot of "big" data, which can then be analyzed in a multitude of ways and turned into meaningful metrics. But with so much data from so many different systems, finding the data that enables improvement can be like finding a needle in a haystack…from multiple barns. Where do you start?

Ankita Gupta
Sr. Analyst,
J.P. Morgan

### *Enter the Cloud!*

First, you need a way to collect and store the data. The cloud has become a preferred option for hosting/storing big data because of the huge storage space availabile and the reduction in infrastructure costs due to third party hosting. The combination of big data, the cloud, and advanced tools has enabled organizations to marry data from different sources, which brings in new perspectives and new opportunities. For example, hotels can use past customer behavior to provide more customized experiences and more targeted offerings, leading to increased loyalty and revenue. This wealth of data, and the tools that help you interpret it, can help you fine tune your organization and ultimately increase your bottom line.

But, every new technology has its challenges and the cloud is no different. While third party hosting is one of the cloud's biggest advantages, it's also one of the biggest risks because the owner is an outsider. Security is of the utmost importance as an organization's data can contain customer credit card numbers, employee social security numbers, and a multitude of additional sensitive information. Loss/interception of this data can have serious legal and ethical issues for individuals and firms, and the resulting bad publicity could follow a company for years.

Is there a formula to do away with these potential security issues? Yes and No.

Risk can never be completely eliminated but it can be mitigated. And to do so, a new version of cryptography, encryption, has come to our rescue!

When encryption is mentioned, many people think about heavy physical servers and keys being shared on CDs and in pen-drives, which needed to be installed to make systems work. Sadly, in this world of increasing costs and extremely large amounts of data, it's impossible to maintain the equivalent infrastructure. Also, most of the primitive formulas don't work too well with the cloud.

So, here's an introduction to some new techniques – Redaction, Obfuscation and On-the-fly encryptions.

### *Types of Encryption*

**Redaction** is the digital form of black-marking the sensitive information on a document (for example, Account no. [redacted]). Redaction can be used to hide directory paths, dataset fields, unique identification numbers, etc.. Products such as Adobe Acrobat Pro DC, Nitro Pro 10, and Rapid Redact are some popular tools used for Data Redaction.

**Obfuscation** replaces the actual data with characters or unmeaningful data so that only authorized people can access it. Remember the xxxxx and ••••••• when you input information like passwords or bank account details? That's obfuscation. This process simply complicates the information enough to eliminate obvious connections or clues to the original data. A simple change of ASCII characters to ANSI, or numbers to their binary forms (ex. 99 ⟶ 1100011), are some basic examples. Therefore, if by some chance the data gets into the wrong hands, there is no logical way of putting the pieces of the puzzle together.

**On-the-fly encryption** is an auto-encryption technique where data gets encrypted automatically every time it is loaded from the server. On-the-fly tools only require basic installation and create a virtual drive that is treated by the system as a typical local drive and any file saved within it is automatically encrypted. The algorithms/mechanism behind the encryption is unknown to the end user. It can encrypt anything from a file to an entire hard disk. It gets decrypted and stored in RAM whenever accessed by users and re-encrypted as soon as it is saved to the server. This is also known as live or transparent encryption, as it happens without user intervention or user knowledge of the encryption/decryption procedure. VeraCrypt, FreeOTFE, DiskCryptor, 7ZIP (for encrypted archive files), and Bitlocker (now pre-installed in Windows) are widely used on-the-fly encryption tools.

In a corporate environment, where the focus is on ensuring data privacy and security with minimum cost, these are the best tools and techniques that can be used.

Be encrypted. Be safe!

**About the Author**

*Ankita Gupta is a Sr. Analyst at J.P. Morgan and a freelance content writer, focusing on analytics. MBA and Btech qualifications, as well as hands-on experience in this industry, have given her a vast insight on the importance of security measures required in a data-driven environment. She has published various articles through major websites like Elance, Skillpages and Crayon Data, one of the fastest growing big data start-ups in Asia. Other than keeping up with the latest trends in IT and information security, she has won several awards as an event manager, is an avid fiction reader, plays guitar and swims for leisure.*