Can we move past 1980 ?

10 August 2018

Roque Martinez

CTO
SmartStream

1980 was a great year. Post Notes had its debut. Cable News Network launched and the US men's hockey team beat the Russian team. Most data was stored in file structures like VSAM. Not until 1983, when IBM launched DB2, did data processing take a leap from file based structures to relational databases.

Between 1980 and today computer processing has grown in speed, cheaper memory, faster storage and most importantly data has become part of the economy. Accessing data has value and getting that data as fast as possible has become normal.

With the advancement of computing, why are market data vendors, exchanges, regulators all providing files to their customers in some form of a file. It really does not matter if the file is json, xml or csv. Files are files and create waste in an environment where the cost of everything is constantly driven downward. I am sure you can on the back of a napkin figure out the cost of processing files, storage of the files, loaders, etl tools, distribution using more etl, etc. Then we need to add the cost of having to acquire the knowledge of how to handle each one of the different file types and vendor types.

Right now there is technology which is making a lot of noise around speculation and investment. Blockchain which is more known for crypto currency is actually the leap reference data needs. Using Blockchain or distributed ledger along with a market utility will remove the waste in processing files.

A distributed ledger allows for all files to cease from crossing network boundaries and facilitate the appropriate timing of data to the right consumers. Imagine having a managed node which contains up to the minute data. Updates are processed when they happen, no waiting to get the end of day file. No more waiting for Saturday to process the full file, to ensure you did not miss an update file during the week.

Users can be provisioned to only access the data which the business line has paid for. Suddenly you have an accurate picture of what data users are requesting. Administrators can report what businesses are using the data. Administrators can stop saying "We can not stop providing attribute X because we have no idea which applications are using the data".

All of this functionality exists, and is waiting to be harnessed to move reference data into 2018. As an industry it will imperative we start moving from our comfortable place in 1980.