

Financial Services and AI: Accelerate Post-Trade Processing

Fast tracking the benefits of Analytics & AI while driving huge operational efficiencies

Is there a [financial services organization in the world who wouldn't like to be at the forefront of harnessing the potential of data and artificial intelligence](#)*?

While AI is the shiny new thing attracting everyone's attention, it should be considered, in the opinion of Financial Services Consultancy Oktay Technology and Pure Storage's own data science team, as simply a more sophisticated set of tools, technologies and techniques within an organization's broader data / analytics capability. That is to say without having a solid, mature, data/analytics capability, organizations will struggle to realize the benefits of AI.

Today, post-trade processing is one of the main "data arteries" for Wall Street firms. While those wise to the potential of data and AI can see the significant transformational opportunities within this huge data set, the reality for many organizations is that this artery is blocked! This is due to a combination of legacy technologies, architectures and siloed business operations that prevent, rather than enable value realization.

CHALLENGE 1 - LEGACY SOFTWARE & DATABASE TECHNOLOGIES

Today the majority of post trade processes are built on old (10+ years) or bespoke, proprietary technologies (e.g. MQ Series, 29 West, Oracle, Sybase) that were not built for the challenges of the modern "big data age". As a result, the architectural constraints of those technologies still to this day dictate the art of the possible for many organizations. An example is how organizations used to have to store data in multiple different relational databases in order to be able to meet the businesses performance and scale requirements, with each database storing a different time range of data:

- Intraday: T+0 intraday
- Short Term: T+1 to 90
- Long Term: T91 - 7 Years

This made it difficult for firms to get a cohesive view of the data, and the further out you wanted to look the longer it would take. Today, with a modern scale-out data storage platform and massively parallel analytic tools like Kafka and Spark, firms are no longer constrained in this way.

CHALLENGE 2 - LUCRATIVE INEFFICIENCY HANGOVER (SILOED OPERATIONS)

A second challenge many firms face is that when a lot of these processes were originally created, FS organizations typically employed what Accenture calls a "lucrative inefficiency" mindset (pre 2008 GFC), giving different parts of the business a great deal of autonomy to build solutions independently to meet their specific business needs. This resulted in a lack of common tools and processes that has complicated

the task of pulling together cross-asset, cross-product views of the business.

Clearly standardizing and consolidating these processes across different assets and products represents a fantastic opportunity to unlock business value with analytics and AI, as well as drive operational efficiencies.

CHALLENGE 3 - LEGACY HARDWARE

While we've talked about the constraints of software technologies used, the same applies to hardware. From a compute perspective it's unlikely that firms will still be utilizing 10+ year old hardware due to Moore's Law. However, from a storage perspective the majority of storage solutions on the market today were not built for the new world of modern scale-out, massively parallel applications like Spark, Kafka and AI etc.

Most storage solutions out there today still have code bases that were written for traditional spinning disk. While retro-fitting solutions with flash provide some performance gains, they fundamentally were not built for these new modern analytic applications, that require what we term "multi-dimensional" performance. This is the ability to provide low latency, high throughput, high IOPS, high metadata performance for any access pattern (serial or random) or file size (large or small) in parallel to 10's, 100's or thousands of hosts concurrently.

Historically, storage systems have been good at delivering one or possibly two of these types of performance, but none have been able to deliver true multi-dimensional performance. This is essentially what Pure's FlashBlade™ array was designed to deliver - high performance for multiple differing workloads at the same time.

In addition to performance, one of the other key things to bear in mind is scale. Today, FlashBlade is able to deliver up to 7.5PB (based on 3:1 data reduction) in a single name space in just 22 rack units.

As a result of these unique performance and scale capabilities, FlashBlade helps customers reduce the number of hardware data silos within the business significantly. In the past, organizations may have had to deploy separate storage arrays for each "application" based on its performance or scale requirements, resulting in inefficiency (multiple purchases, arrays to manage, duplicated data etc). Today with FlashBlade, depending on your organization's volume of data, you can potentially accommodate the full E2E post-trade process on a single FlashBlade system.

Furthermore, as performance on FlashBlade is so abundant, there is no need to copy data to a separate high-performance storage array in order to undertake computationally intensive tasks (such as AI for example). This drives even greater efficiency and agility (as large data sets don't have to be copied over the network from A to B) etc. This is what we call our "[data hub](#)" strategy**.

CHALLENGE 4 - DISAGGREGATING STORAGE & COMPUTE

Depending on when an organization last modernized its post-trade processing, it may either still be using traditional relational databases or may have made the transition to some more modern database/analytic architectures (e.g. Hadoop, Spark, Cassandra, Mongo etc). These more modern architectures typically leverage their own bespoke distributed direct-attached storage implementations across multiple storage nodes. Whilst DDAS architectures help solve the scale challenge, they typically struggle with operational efficiency (fundamentally the need to scale compute and storage in unison does not make sense when 99% of organizations do not have performance requirements and capacity requirements in lock step) and performance at scale.

Increasingly there is an emerging consensus among analysts, application vendors and customers that tomorrow's architectures should be disaggregated. One of the major reasons for this is that disaggregated storage and compute is a prerequisite to be able to dynamically scale compute resources – one of the key benefits of public cloud and containerization – both “hot” technology areas that the analytic tool vendors are keen to leverage. Great examples of this include Splunk's move toward an S3 backend for their new SmartStore architecture**** and Cloudera/Hortonworks increased support of S3.

FlashBlade allows clients to disaggregate storage and compute on premise with far greater performance (S3 on Pure is 10X faster time to first byte than AWS and is 100X faster indexing 100 objects). This performance is critical to realizing the goal of having a single “data hub” underpinning all post-trade processes and avoiding the proliferation of multiple different data silos and duplicate data that exists today – all of which ultimately leads to higher costs, lower operational efficiency and slower time to result.

SUMMARY - MODERNIZE POST-TRADE PROCESSING

The opportunity for Wall Street firms is to modernize post-trade processing by:

- Removing bespoke/proprietary/legacy technology from the post trade process
- Consolidating, standardizing and streamlining data processes
- Consolidating and standardizing tooling, leveraging new open source analytics tools like Spark and Kafka that are built for massive scale and parallelism
- Standardizing and simplifying real time, cross asset, cross product views of the business

The benefits:

- It unlocks the value of Analytics and Artificial Intelligence for your business
 - By having high quality, consolidated, integrated post-trade data
 - By providing a consistent **real time** view of the business, allowing better deployment of capital, more accurate view of risk etc.
- Reduce cost – through consolidation, standardization and removing duplicated processes and data

*by artificial intelligence we are referring to principally machine learning and deep learning rather than the creation of chatbots etc.

** = for more information on our [data hub](#) strategy please read the following: