XONA PARTNERS

Data Sciences

Applications in Financial Technologies

Dr. Riad Hartani,
Dr. James Shanahan (Xona Partners)
Paddy Ramanathan (Digital Confluence)

January 2014

Applied Data Sciences in Financial Technologies

The 1980s represented a transformational period for banks where Information Technology systems virtually transformed how banks conducted their business. The 1990s was the era of online banking where we saw a dramatic shift in how banks sold and serviced their customers. The 2000s represented a gradual maturity and shift towards the multi-channel banking. Today, banks have the opportunity to reinvent themselves leveraging upcoming IT transformation towards Big Data architectures, and leveraging data sciences techniques for various analytics applications, with a direct impact in all aspects of their business from customer engagement, operational efficiency and risk management. All revenue, operational efficiency and risk management decision making in a bank can be infused and optimized with Big Data analytics.

We have recently been analyzing applications of novel data sciences as a collection of advanced intelligent data analysis techniques, such as evolution of machine learning and data mining models to practical industry problems, with various experimentations to specific problems in select industry verticals. This is timely and opportunistic, given that data science has made the big leap from being a research topic to a set of tools accessible in various shapes and forms to different industry verticals and optimized to resolve some of their more challenging problems.

In this paper, we synthesize our experience on the experimental front through a recent case study, applying and customizing select advanced data science algorithms to a new set of financial services and technologies applications. Specifically, we address the problematic of risk management, fraud detection, financial customer characterization and framing of new financial services offers to various customer segments. Specifically, we highlight a set of problems and use cases that we have been experimenting with and developing solutions for in the financial technologies realm.

Data Management in the Financial World - Snapshot

To achieve the stated goals above, we opportunistically leverage the fact that few concurrent trends are converging, when it comes to data management overall and the specifics seen in the financial world. A brief snapshot is presented here.

First is the maturity of data management models

We are witnessing the fast adoption of novel architecture to store and access large data sets (Hadoop, MapReduce, HDFS, Yarn, etc. – commonly known as Big Data models), as well the commercial availability of various cloud deployment architectures (OpenStack, vCloud, Cloudstack, AWS, etc.). This is removing significant logistical obstacles to embracing management of large data structures. The move is likely to be even more significant moving forward, given the immense number of contributions of the open source community in this area. Key here is convergence onto universally adopted platforms versus what was before seen as a proliferation of diverse platforms.

Second is the evolution of data sciences

This applies to the large set of data analysis models in a broad sense, and specifically machine learning and mining algorithms that are more accurate and computationally tractable, leveraging distributed cloud-based computing models. Current developments in Deep Learning, for example, illustrate how an older field of neural networks achieved breakthroughs in accuracy when its

algorithm improvements were fueled by much increased computational power. Taking advantage of the introduction of new computing models, such as algorithms parallelization, GPUs and alike, then porting that to distributed cloud compute models, not only the existing algorithms have been optimized to run better and faster, but a number of additions and optimization have been developed and run in a computationally tractable way.

Third is data availability

Leveraging compute and storage architectures that are increasingly scalable to selectively and dynamically process large volumes of data, relying on various models of data capture, via sensors, devices, and management interfaces. Large data sets influence algorithm choices by easing the risks of over-fitting, which leads to better generalizable insights. The sheer size of data available is likely to increase, either as front-end data in real time or backend data stored as historical patterns. In the financial world specifically, data collection architectures have evolved in a way that allows for data to be captured fast enough for deeper analysis, and software-based data management architectures in a way that data can be queried, received and presented to relevant data processing models.

Leveraging ongoing IT transformation initiatives

Along with the trends listed above, one additional specific reason why the opportunity presents itself to leverage specific data science technologies in the financial applications world, is that the lead financial institutions are, right now, in the process of putting in place specific data management and IT transformation models. This in turn addresses aspects of data availability, presentation, scalability and reliability. A brief description is provided here.

Like various information technology players, financial institutions are, or will soon be considering enhancing their data infrastructure to help build data solutions which will optimize their ability in identifying, capturing, and managing data to provide actionable, trusted insights that improve strategic and operational decision making, resulting in incremental revenues and a better customer experience. The current challenges of the existing platforms mostly affect the operations teams' ability to provide reliable SLAs for the IT execution processes that are critical for the business, the scalability of the data platforms, to perform effectively as the business data volume and messaging grows. As such, the desired goal is to create a solid foundation architecture that is able to provide these optimal functional capabilities, and a platform to overlay additional applications such as intelligent business intelligence and data science as a service capability.

As of today, the existing information infrastructure and analytics processes suffer from challenges we have observed and worked on in the most typical large scale data and IT projects, including those in banks and financial institutions, some of which are listed below, and IT transformation projects' main focus is on providing solutions to these challenges:

- Data needed in past or present time continues to change for several days due to intake problems
- Stringent SLAs are not met for business critical jobs
- Implementation and orchestration of software jobs is fairly unproductive

- Incomplete metering, monitoring, and diagnostics of distributed software jobs and processes
- Increasing time of completion for software jobs as ad-hoc usage increases
- Data architecture issues regarding naming, change management, and running analytics
- Lack of comprehensive policy-based management for retention, replication, archival, and compression
- Real time run Analysis, only generates reports and events from near real-time data
- Provide near-real time granularity for generating comprehensive reports as opposed hours or days
- Consolidate multiple data centers to a combined cluster for data analysis.

These data architectures are fast evolving and our contributions have focused on addressing the various challenges above, which are common to most financial institutions' IT backend and data management architectures.

Financial Application Optimization Case Study

Within the realm of financial applications, and based on where banks and financial organizations are focusing their business priorities today, some of the key areas we see in the short to medium term having the biggest wins with Big Data analytics are:

1) Risk management – credit risk, payment fraud detection/AML
2) Sales and marketing – 360-degree view of the customer, anticipating customer financial needs
3) Product innovation through creative combination with digital enablers like mobile

We address the first use case in detail in this section, and provide brief highlight of the analysis work we have been pursuing for the other use cases.

Case Study 1: Risk Management –Payment Fraud Detection

According to IDC research, risk technology spending in banks is to increase at a compounded annual rate of 6.9% until 2017. This spend comes at the tail of heavy investments in risk technology in the last 5 years. The spending is towards compliance systems, counter fraud, credit and cyber-protection and information security.

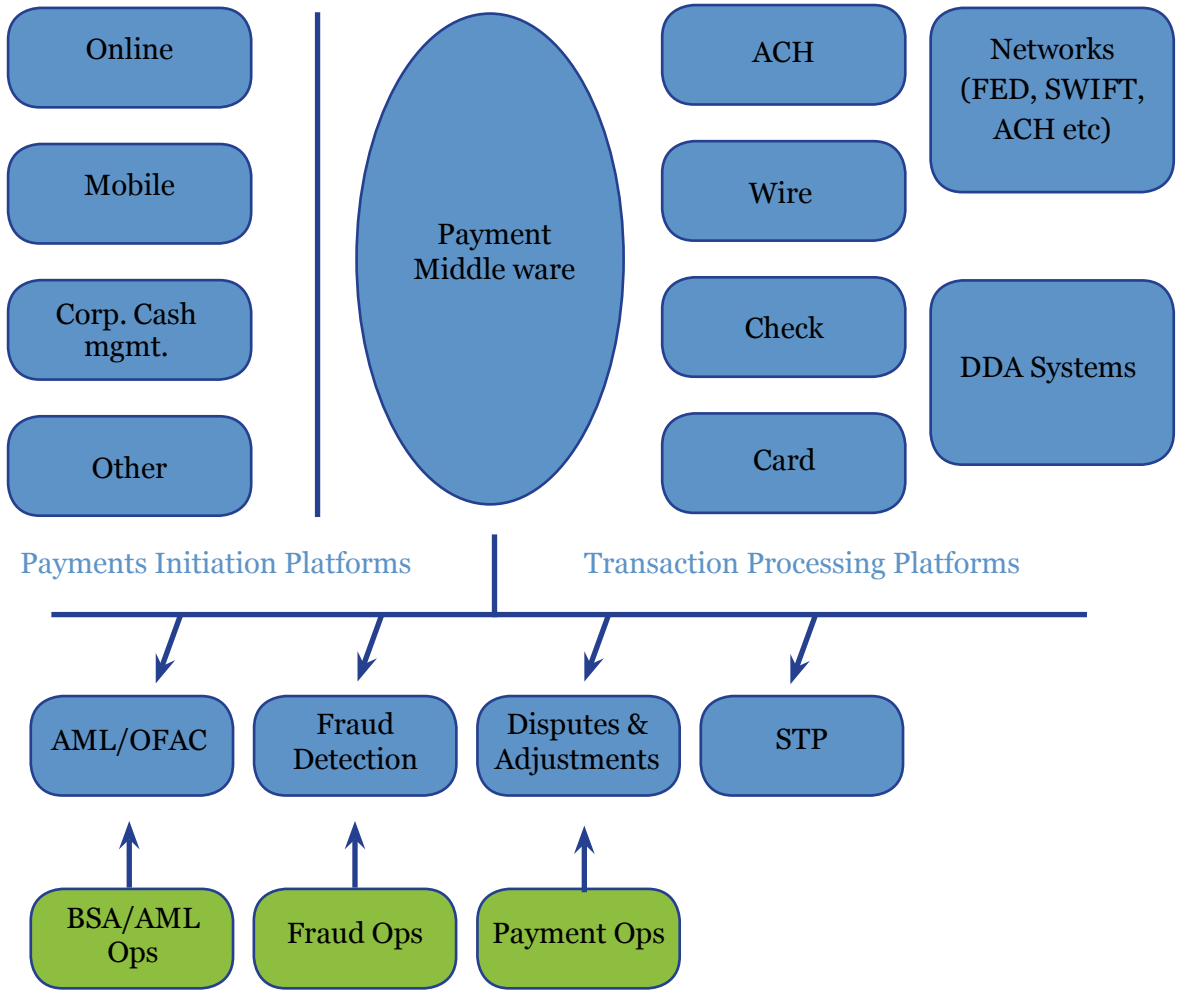
Big Data Analytics has the potential to disrupt the way financial institutions manage their risk. The quantum leap in scale, affordability and the variety of analytics presents an opportunity to increase the effectiveness of risk by getting more accurate prediction of total and residual risk, risk trends and timeliness of alerting, and the overall cost of managing risk.

Fraud Detection in Payments – Context

Included in our considerations are aspects such as domestic and international wires, ACH, non-urgent international payments.

The total amount of wire transfers in exceeds US$130 million transactions and the dollar value of all transaction is about US$600 trillion. These numbers have more than tripled in the last 20 years making it very attractive to fraudsters.

Today banks use fraud analytics extensively. Every payment is scored to ensure the integrity of the initiator and validating the intent of transaction. Current payment processing infrastructure includes a payment middleware (such as a Payment Hub) and AML and Fraud Detection from vendors such as Actimize. The Fraud team reviews transactions that the analytics identify as high risk. A relatively large amount of false positives are generated that require manual work to clear the payment.



The above figure represents a typical payment ecosystem at a financial institution. The risk and regulatory functions are modularized in the AML/OFAC and Fraud Detection capability, which is implemented to varying degrees of sophistication depending on the institution. Typical vendor products used by the institution varies depending on the size and complexity and existing relationship but some of the common vendors include Actimize, FICO, ACI Worldwide, SAS, etc.

The financial impact of fraud for a financial institution is more than the actual worth of the fraudulent transaction. It includes brand and reputational impact, fines, customer churn, investigation cost and remediation costs.

Fraud happens when the credentials are compromised. There are several options for a fraudster to acquire credentials besides just by accident: some of the common schemes are malware, social engineering, phishing, fishing, and email compromise.

Current State of the Art

The following techniques are in use in detecting fraud:

- Calculation of statistical parameters (e.g., averages, standard deviations, high/low values) – to identify outliers that could indicate fraud.
- Classification – to find patterns amongst data elements. Stratification of numbers – to identify unusual (i.e., excessively high or low) entries.
- Digital analysis using Benford's Law – to identify unexpected occurrences of digits in naturally occurring data sets.
- Joining different diverse sources – to identify matching values (such as names, addresses, and account numbers) where they shouldn't exist.
- Duplicate testing – to identify duplicate transactions such as payments, claims, or expense report items.
- Gap testing – to identify missing values in sequential data where there should be none.
- Summing of numeric values – to identify control totals that may have been falsified.

Validating entry dates – to identify suspicious or inappropriate times for postings or data entry.

Current Challenges

Based on this specific use case, and the underlying data sets, software in use today, we have observed the following challenges and limitations:

- Each vendor has its strengths but there is limited standardization. Setup, configuration and maintenance are expensive functions. A typical project to implementation can take over 6 months
- Limited satisfaction of customers even when most up to date solutions are applied
- There is a huge percentage of false positives increasing the cost of fraud operations
- Limited link analysis and social network data integration
- Limited visualization capabilities of customer behavior and outliers
- Limited behavior analytics based on individual, segment, population segment models
- Multi-channel integration requires custom integration of the various channels like mobile and online
- Limited outlier detection and analysis with a variety of data like log files, textual analytics in transaction records.

Data Science Contributions

Based on this specific use case, and the underlying data sets and models in use today, we have been working on putting together an architecture, design and implementation of some of these algorithms with the goal of making fraud or AML detection in payments more efficient (infrastructure and labor cost of support), requiring less setup (self-learning), and adaptable to changing landscape of payments. Analysis goals aim at reducing false positives, increasing response times for detection and reducing the cost of setup and support of these capabilities in the banks. Mobile payments (even wires initiated or approved from mobile devices) are the fastest growing payment initiation point for commercial payments (i.e. company to company payments).

The data science solutions we have been working on to address some of the use cases highlighted above are based on the two engineering approaches below:

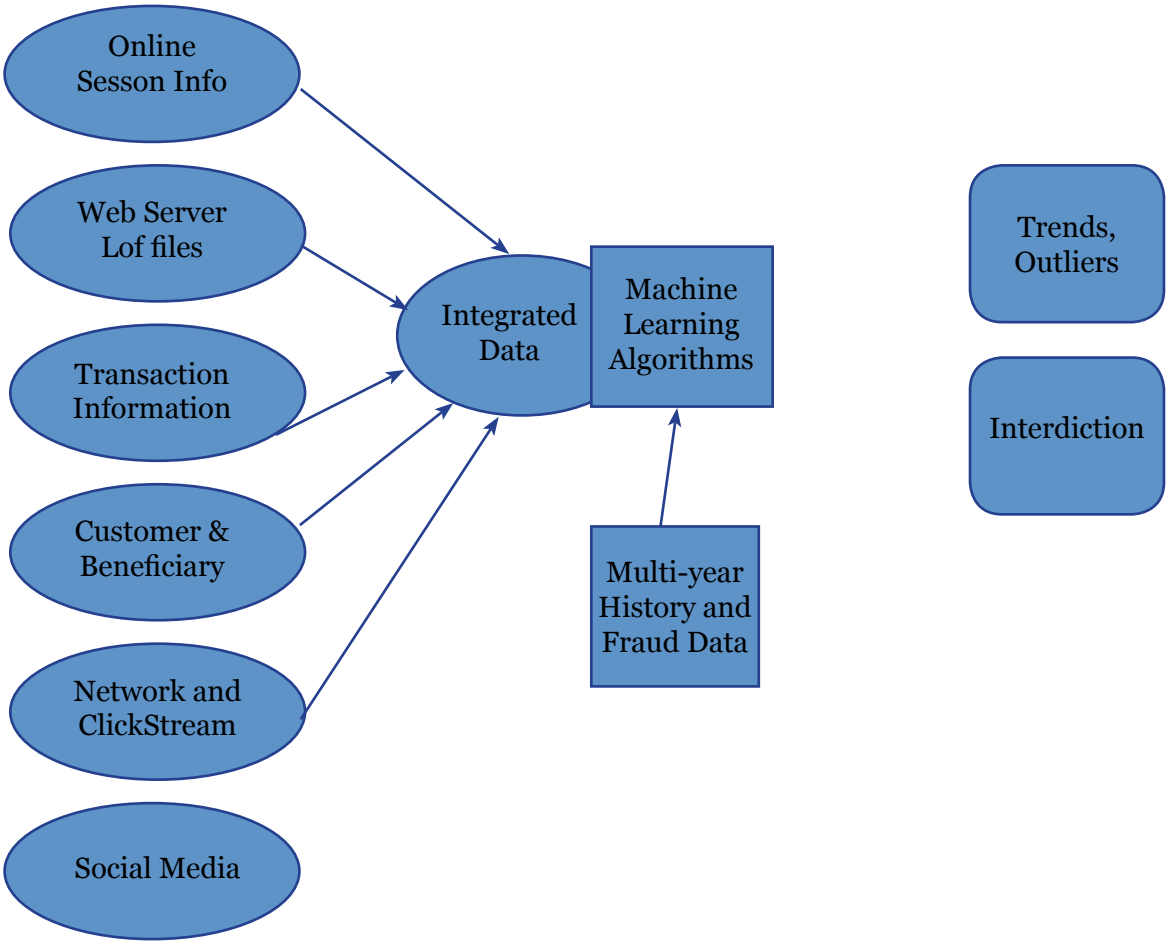
(1) A hybrid local/cloud based data gathering and storage solution leveraging novel techniques optimized for a variety of data models. Adaptations of Hadoop-like models and their underlying MapReduce computing paradigm for large-scale distributed file systems are leveraged to present the various data sets, that normally gather in silos, into a common data representation accessible to data processing models.

(2) A set of machine learning and data mining algorithms specifically focused on clustering and predictive modeling in high dimensional spaces based on imprecise, uncertain and incomplete information, efficient statistical data summarization and features extraction algorithms as well as large scale real time data streams management. These tools will be at the core of the processing engine, and will aim at deriving optimization to the existing business logic and augment it with appropriate business process logic, which would be mapped to a set of new revenue generating services.

The results foreseen from the analysis we are engaged in, addresses some of the challenges highlighted above. A specific focus is performing correlation and link analysis in real time and present the results in a visual, and through that, derive pertinent connections between enterprise, online and social media information in real time and present in a visual manner to reveal relationships for detection and investigation. Things that we believe are useful to identify fraud rings and improve detection accuracy.

The two other angles we have been focusing our efforts on are briefly described below. The focus, as for the fraud detection problem has focused on analyzing state of the art, understanding the data structures and exiting models in use today, zooming in on the business challenges and leveraging various data science techniques to improve upon such challenges.

A conceptual Big Data Fraud Detection Engine is highlighted below:



There are ancillary benefits of such architecture as the data can be used for gathering other insights for marketing or service improvement initiatives.

Case Study 2: Sales and Marketing - 360 Degree of Customer

Big Data and associated visualization techniques have the power to integrate customer relationships, interactions with the enterprise with the appropriate context for a wide variety of audience to provide a rich experience for the customer. The relevancy of the information is critical both to the context of the interaction and the overall context of the relationship with the customer.

Case Study 3: Product Innovation – Card Linked Offers

This addresses the angle of digital natives vs. non-digital natives scenarios. Banks ability to monetize their data and offer value-added services to their customers is critical in this digital age. Big Data provides a way for banks to not only improve their own decision-making but also provide value-added services to their customer to deepen their relationships and increase their loyalty. Card-linked offers is a data driven mechanism where banks can information in transaction data to provide custom offers and deals to their customers.

Conclusions

A brief description of some of the data science applications to financial applications has been highlighted, as a way to demonstrate applicability and value of such techniques in the real world. Specifically, one demonstrates that financial organizations, which have historically been fairly slow moving in terms of pushing new data analysis techniques, are starting to get disrupted. Disruption in this case is beneficial, as it will likely converge on making operations way more efficient, successfully catch fraud and reduce the total cost of ownership of technology, taking full advantage of the potential of data science models.

Our team, with its diverse technology expertise in IT transformation towards Big Data architectures as well as data sciences, in conjunction with financial technologies and businesses specific knowhow, has been working with select players, in a win-win model, to solve some of the leading multinational pain points – or allow them to develop an edge in what is sure to be an increasingly competitive and apt for disruption market place.

Xona Partners (Xona) is a boutique advisory services firm specialized in technology, media and telecommunications. Xona was founded in 2012 by a team of seasoned technologists and startup founders, managing directors in global ventures, and investment advisors. Drawing on its founders' cross functional expertise, Xona offers a unique multi-disciplinary integrative technology and investment advisory service to private equity and venture funds, technology corporations, as well as regulators and public sector organizations. We help our clients in pre-investment due diligence, post investment life-cycle management, and strategic technology management to develop new sources of revenue. The firm operates out of four regional hubs which include San Francisco, Paris, Dubai, and Singapore.

Xona Partners

www.xonapartners.com

advisors@xonapartners.com