insideBIGDATA

insideBIGDATA Guide to

Big Data for Finance

By Daniel D. Gutierrez

Photo: everything possible / Shutterstock.com

BROUGHT TO YOU BY DELL Technologies AMD EPYC

Contents

Introduction 2
Retail Banking 3
Regulatory and Compliance 5
Algorithmic Trading 6
Security Considerations 7
Conclusion 9
Next Steps with Dell Technologies and AMD 10

Introduction

Across industries, data continues to grow as an extremely valuable resource. This is especially true in the financial services sector. "Financial services" is a broad term that encompasses commercial banks, investment houses and insurance companies. This highly competitive sector has been largely dominated by global conglomerates, with a diverse range of smaller companies.

Financial services institutions (FSIs) have always been a substantial consumer of information technology, usually ahead of other industries.[1] For the past several years, FSIs have continued that trend by leveraging big data analytics and artificial intelligence (AI) to enable new opportunities and deliver benefits to customers and employees alike.

About a decade ago, big data became an emergent trend driving investments in enterprise analytics, and correspondingly, analytic excellence is central to much needed innovation in today's financial services marketplace. Business analytics applied to capital management, regulatory compliance, corporate performance, trade execution, security, fraud management, and other instrumental disciplines is the principal innovation platform to improving strategic decision making.

But now, the financial services sector is witnessing an era of digital disruption and innovation, driven by the following macro trends:

- **Increased regulation:** In the aftermath of the 2008 financial crisis, financial service institutions (FSIs) have become subject to greatly increased regulations for controlling risk and understanding exposure. Risk controls previously done weekly or nightly are now done several times per day — sometimes in near real time. Compliance with regulations, such as the Fundamental Review of Trading Book (FRTB) and the Comprehensive Capital Analysis and Review (CCAR), involve simulating a huge number of parallel market scenarios and over large portfolios simultaneously. This requires computing power that stretches the limits of traditional computing.
- **High consumer expectations:** Tech savvy consumers are increasingly seeking financial services that are faster, cheaper, more personalized and easier to access.

1 Mckinsey 2018; Naveira, Jacob, Rifai, Sinmon, & Windhagen, 2018.

- **New, more agile players in the market:** A growing field of financial technology (fintech) companies is shaking up traditional business models, eroding market share and undercutting margins for incumbents.
- **Emerging technologies:** Breakthrough technologies such as artificial intelligence (AI), including deep learning and machine learning, are allowing early adopters — both fintechs and established companies — to gain massive differentiation and will produce significant disruption over the coming years.

With mounting regulations, customer expectations and disruptors around every corner, FSIs will need to embrace digital transformation to prosper and stay competitive. FSIs recognize the power of technology to shape their futures, and CIOs now drive a large part of the company strategy. This new generation of FSI IT leaders are seeking to securely and rapidly enhance their digital capabilities to reduce costs, grow share and provide better customer experiences.

The new generation of FSI IT leaders are seeking to securely and rapidly enhance their digital capabilities to reduce costs, grow share and provide better customer experiences.

The goal for this Guide is to provide direction for enterprise thought leaders on ways of leveraging big `data technologies in support of analytics proficiencies designed to work more independently and effectively across a few distinct areas in today's FSI climate:

- Retail Banking
- Regulatory and Compliance
- Algorithmic Trading
- Security Considerations

Retail Banking

Thanks in large part to the availability of data and the movement from in-person to online banking, today's banking institutions look very different than those of just a decade ago. As the scale of data is overwhelming traditional systems, banks must adapt to new technologies to unlock the power of their data.

Banks have long been considered innovators when using data analytics to tackle numerous business challenges such as risk management, fraud detection and price discovery. Yet today, the volume of data is much bigger and more diverse than ever before.

At the same time, regulatory demands for banks have become much more stringent and the increased level of historical storage requirements have meant that banks must keep data for longer periods of time. This complexity has necessitated new approaches and technologies leading to big data infrastructure.

There are a number of motivating factors for engaging big data technology in support of retail banking:

- The availability and scale of data is extraordinary and requires a new technology mindset.
- A significant transition from in-person to online banking resulting from the ease and affordability of executing financial transactions.
- A distinct need to collect and analyze this information in order to accurately assess risk and market trends.
- The availability of new data sources such as data from social media, blogs and other news feeds offer significant new opportunities. As with all online markets, banking is competitive and banks are interested in using any opportunity, identified through data, to cross sell and upsell customers.

Coupled with the above motivations, there are a number of potential application areas for big data in retail banking:

- Since banks weigh the risk of opening new accounts versus the opportunity to hold deposits, big data can screen new account applications for risk of default. In addition, big data can identify high-risk borrowers for auto loans.
- Maximizing customer touch points leading to better customer service and churn detection and avoidance.
- Cross-sell/up-sell using recommender systems.
- Monetize anonymous banking data in secondary markets.

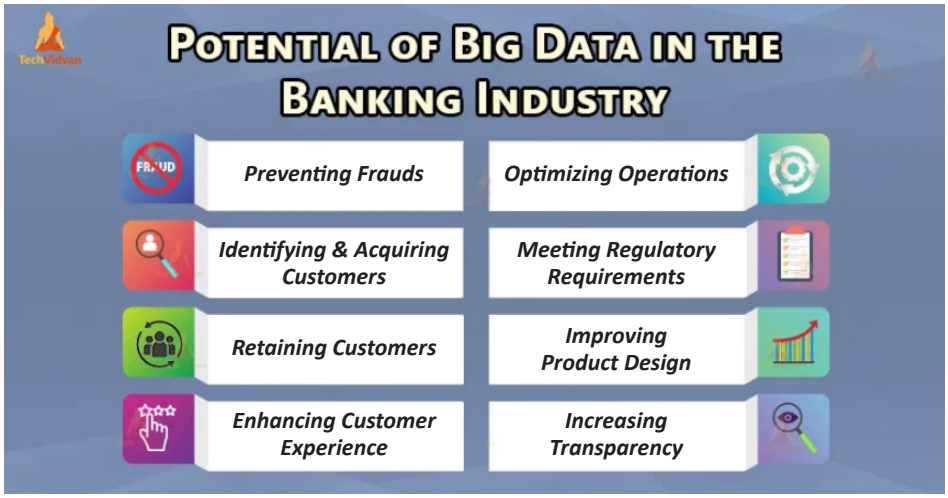
Solving the challenging problem of customer churn is one good example of how big data can make a real difference for retail banking. With credit card transactions, loyalty programs, and many other customer touch points, banks possess far more data about customers than any other industry, but in spite of all the data, customer-centric companies like banks are often unable to deliver effective personalized service. The main reason is the low level of customer intelligence.

The world is increasingly interconnected, instrumented and intelligent and in this new world the volume, velocity, and variety of data being collected is unprecedented. As the amount of data created about a consumer is growing, the percentage of data that banks can process is going down fast. Based on their engagement with popular consumer destinations like Amazon or Yelp, customers have expectations about similar experiences from the banking applications. Without deep know-how about their customers, banks may not be able to meet these expectations.

The result? Lost revenue opportunities, low coupon redemption rates, lower share of customer's wallet and lost competitive agility. In a nutshell, not being able to gain insights from the goldmine of data means banks are allowing their competitors to identify critical business trends and act on those before they can, ultimately losing business.

In summary, in order to advance the level of customer intelligence banks must:

- Leverage big data to get a 360-degree view of each customer.
- Drive revenues with one-to-one targeting and personalized offers in real-time.
- Reduce business risk by leveraging predictive analytics for detecting fraud.
- Achieve greater customer loyalty with personalized retention offers.
- Employ the power of big data without worrying about complexities and steep learning curves.



Source: TechVidan

There are many quality software tools allowing banking institutions to reap the benefits of big data. For example, many FSIs use Splunk to collect information as an industry leading big data platform designed to integrate information of all types into easily deployed visualizations.

HOW DELL TECHNOLOGIES HELPS

Dell Technologies and Splunk partner to make adopting Splunk simpler, by engineering a portfolio of purpose built solutions with non disruptive scalability and performance optimized for Splunk workloads. Together, Dell Technologies and Splunk enable you to harness the power of machine data analytics with simplified deployment and scalability.

In summary, the banking market and consumers who utilize finance products generate an enormous amount of data on a daily basis. While each activity is a single data point, multiple pieces of information creates a larger picture that can be used to recognize patterns in customer behavior, purchasing choices and other key insights. AI and data analytics have changed the way all this information is processed, making it possible to identify trends and patterns which can then be used to inform business decisions, in real time, at scale. Leveraging technologies like Splunk and Dell Technologies, help FSIs achieve this desired outcome.

Regulatory and Compliance

It is important for banks, investment firms, and other financial services organizations to be able to collect and analyze this information in order to accurately assess risk and determine market trends. This became apparent during the market downturn of 2007-2008, when banks and brokerage houses scrambled to understand the implications of massive capital leverage and their ability to model and refine liquidity management.

A single bank might capture internal transactions exceeding two billion per month, in addition to collecting public data of over a billion monthly transactions. These tremendous transaction volumes have made it nearly impossible to create models that take into account multi-year data sets using detailed data.

Financial firms manage anywhere from tens to thousands of petabytes of data, yet most systems used today build models using only samples as small as 100 gigabytes. Relying on data samples requires aggregations and assumptions, resulting in inaccuracies in projections, limited visibility into actual risk exposure, instances of undetected fraud, and poorer performance in the market. As result of more rigorous regulatory compliance laws, the financial services industry has had to

Regulatory Big Data. Source: Moody's Analytics

store an increasing amount of historical data. New technology tools and strategies are needed to address these demands.

Hadoop represents a good path for financial sector firms to adopt big data. With Hadoop, firms have access to a powerful platform providing both highly scalable and low cost data storage tightly integrated with scalable processing. Financial firms are now able to tackle increasingly complex problems by unlocking the power of their data. The capability to understand and act upon their data opens the door to a richer and more robust financial ecosystem.

Spark is an open-source data analytics cluster computing framework built on top of HDFS. Spark serves as evidence of the continuing evolution within the Hadoop community — away from being a batch processing framework tied to the two-stage MapReduce paradigm to a more advanced in-memory, real-time platform. Now, FSIs can better serve their customers, understand their risk exposure and reduce incidents of fraud.

Hadoop Benefits

Here is a short list of benefits afforded by Hadoop to the financial services industry:

- ✓ Scalability to capture and analyze data previously untapped
- ✓ An economical way to store and process data
- ✓ A central repository for various data structures from existing and new data sources
- ✓ Operational efficiency by moving jobs to technology designed to process multiple data types
- ✓ The ability to ask different questions to improve decision making

Mastercard applies 1.9 million rules to 165 million transactions per hour in a matter of milliseconds.

HOW DELL TECHNOLOGIES HELPS

Dell Technologies has invested to create a portfolio of Ready Solutions designed to simplify the configuration, deployment and management of Hadoop clusters. These trusted designs have been optimized, tested and tuned for a variety of key Hadoop use cases. They include the servers, storage, networking, software and services that have been proven in our labs and in customer deployments to meet workload requirements and customer outcomes.

The modular solution building blocks provide a customized yet validated approach for deploying new clusters and scaling or upgrading existing environments. Ready Solutions for Hadoop have been jointly engineered to optimize investments, reduce costs and deliver outstanding performance.

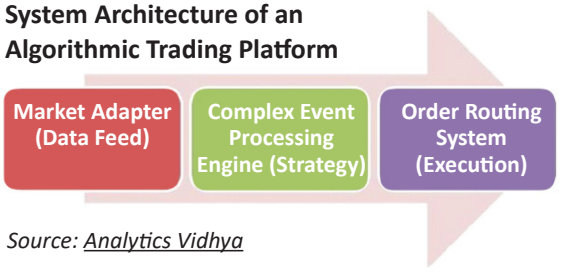
Algorithmic Trading

In the digital economy, data — and the IT solutions used to harness it — are often a financial services company's prime source of competitive advantage, as more automated the process, the faster the time to value. This is especially true for algorithmic trading, a highly automated investment process where humans train powerful software applications to select investments and implement trades automatically.

The ultimate evolution of algorithmic trading is high frequency trading, where the algorithms make split second trading decisions designed to maximize financial returns. Automating and removing humans from trading has several advantages, such as reduced costs and greater speed and accuracy.

Developing trading algorithms requires a proprietary mix of data science, statistics, risk analysis and DevOps. Then the algorithm is back tested, which involves running it against historical data and refining the algorithm until it produces the desired profits. The algorithm is then put into production, making trades in real time on behalf of the firm. The real world yields produced by the algorithm produce even more data, which is used to continually train the algorithm in the back end and improve its performance. This training feedback loop is a data intensive process.

System Architecture of an Algorithmic Trading Platform



Source: Analytics Vidhya

More recently, developers have taken up machine learning, a subset of artificial intelligence (AI), to improve predictive capabilities, using deep neural networks to find trends that trigger buy or sell decisions. In addition to automation and intelligence, high frequency trading platforms deliver competitive advantage by placing thousands of trades before the market can react. Therefore, high frequency trading has led to competition in computational speed, automated decision making, and even connectivity to the execution venue to shave off microseconds and beat other traders to opportunities.

What's more, financial trading firms are continually developing, implementing and perfecting algorithmic trading strategies to stay a step ahead of the competition. This puts significant stress on infrastructure because the algorithm must continuously adapt to new input to remain relevant. As such, the back end infrastructure must accommodate for live data feed and quick processing of large amounts of data. Databases must be able to feed the compute engine in real or near real time to update the algorithm.

The data intensive training requirements and the need for high speed and low latency mean that these sophisticated algorithms are typically trained and run on High-Performance Computing (HPC) systems to provide the rapidity and accuracy required to dominate the market. A HPC system that supports algorithmic trading should be able to accommodate current workloads seamlessly and provide the flexibility, performance and scaling required to continually train and update algorithms to stay ahead of the market.

HOW DELL TECHNOLOGIES HELPS

Dell Technologies has the expertise and experience to design and implement HPC, data analytics and AI solutions optimized for algorithmic trading. This includes considerations for software, services and infrastructure design with complete architectural design examples, such as:

- **Data lake configurations for data ingestion using streaming tools such as Apache® Kafka® and StreamSets® aimed for low latency real time data feed with the Ready Solution for Data Analytics Real Time Data Streaming.**
- **Apache Hadoop® with Cloudera® and Greenplum® supported by Dell EMC Ready Solutions for Hadoop.**
- **Dell EMC Ready Solutions for Data Analytics with Spark® on Kubernetes and Data Science and Advanced Analytics with VMware Tanzu.**

Security Considerations

An important application area where big data is taking a firm foothold with many FSIs is information security. In conjunction with the traditional 3 Vs of big data, financial industry firms must consider a fourth V: **vulnerability**. To manage big data effectively, you must keep it secure and compliant with regulatory requirements at all times (vulnerability).

Protecting a vast and growing volume of critical information—and being able to search and analyze it to detect potential threats—is more essential than ever. As the software platforms (e.g. Hadoop) supporting this quantity of data are more widely used, managing their security and availability becomes a big data challenge in and of itself, requiring continuous diagnostics and monitoring.

Banking and financial institutions need to secure the storage, transit and use of corporate and personal data across business applications, including online banking and electronic communications of sensitive information and documents.

The typical IT environment consists of a mix of new and legacy systems and applications across highly distributed networks of branch offices, call centers and web portals. Many of the traditional point security solutions that are deployed add complexity and management costs, and leave gaps between systems and applications that are highly vulnerable to attack.

The increasingly global nature of the financial services industry makes it necessary to comprehensively address international data security and privacy regulations. Financial institutions are top targets of cybercrime. While all types of businesses are vulnerable to attacks by criminals, it's the security breaches at financial firms that elicit the most media attention, public scrutiny and legislator consternation. When threats occur, it's more than financial loss at stake.

In today's business world, cybercriminals are lurking everywhere — looking for new ways to gain unauthorized access to sensitive data. Some savvy criminals even use AI in their exploits, so your company has to use AI to fight AI, or risk leaving your critical assets exposed. Security has to be prioritized at every layer of your IT infrastructure, especially as AI, the Internet of Things (IoT), and other evolving technologies introduce new vulnerabilities.

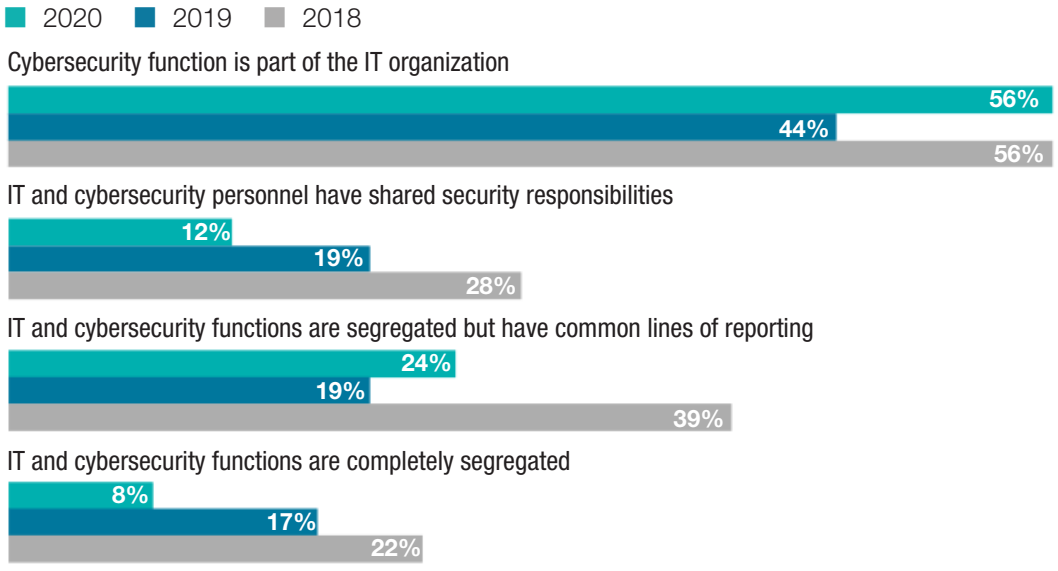
Data privacy, digital ethics, and security best practices are vital for maintaining trust and ensuring business success. To protect against emerging threats, your infrastructure should be AI-ready with end-to-end security.

At the platform level, your systems should include built-in security with:

- A secure hardware root-of-trust. If the server BIOS is corrupted or compromised, the processor should prevent the firmware from booting.
- Hardware-accelerated memory encryption. The processor should protect data-in-use with full-stack encryption, using secure key generation and key management.
- Virtual machine (VM)-level encryption. The processor should also protect against hypervisor corruption with hardware protection, which is a more robust solution than software protection.
- Server-level security technology. Beyond secure processors, the server platform should offer digitally signed firmware updates, automatic BIOS recovery, firmware rollbacks, and systems lockdowns to prevent configuration "drift."
- Trusted vendors also build security into their manufacturing processes and ensure a secure supply chain. This includes verifying the authenticity of components or parts, using trusted suppliers, and physically securing the build environment, the system build process, and the final shipments to customers.

More than half of large financial respondents from large firms had cybersecurity as a part of their IT organization

Cybersecurity integration with IT for large FSI respondents



Sources: FS-ISAC/Deloitte Cyber & Strategic Risk Services CISO survey reports, 2018, 2019, and 2020; Deloitte Center for Financial Services analysis

Above the hardware platform requires intelligence against cyber threats as well. Dell EMC PowerProtect Cyber Recovery provides maximum control of data and infrastructure with a secure on-premises vault protected with an operation air gap and multiple layers of physical and logical security.

Dell EMC PowerProtect Cyber Recovery provides proven, modern and intelligent protection to isolate critical data, identify suspicious activity and accelerate data recovery allowing you to quickly resume normal business operations.

- ✓ On-premises in an air-gapped secure Cyber Recovery vault
- ✓ Compliance-level hardware-based immutability and NTP tamper protection
- ✓ CyberSense identifies threats and helps enable assured recovery

Dell EMC PowerProtect Cyber Recovery provides proven, modern and intelligent protection to isolate critical data, identify suspicious activity and accelerate data recovery allowing you to quickly resume normal business operations.

Protecting your vital data from cyber-attacks requires proven and modern solutions. With Dell EMC PowerEdge Servers, using the latest AMD EPYC Processors, as well as Dell EMC PowerProtect Cyber Recovery, customers can feel confident in identifying and restoring known good data to resume normal business operations after a cyber-attack.

Conclusion

The financial services industry was quick to adopt data analytics and AI, even in its early stages. One big and exciting frontier for AI and analytics in this industry is the ability to combine both structured and unstructured data to continuously gain new insights, accelerating data to value. Whether it's retail banking continuing to drive more digital-first initiatives, companies keeping compliant with increasing data growth, while keeping it secure and cost-efficient, or bringing in more automation, AI and ML into trading strategies and processes, FSIs that can create the most complex real time data models to incorporate the most data, will gain huge differentiation in their businesses.

To avoid missteps, it also helps to have the right technology partners to help get you there.

A race is already developing where FSIs are competing to incorporate the richest collection data into their AI models to gain the deepest insights. These budding super users will be the new AI enabled tech stars of finance, much as Google®, Amazon®, Facebook®, Baidu® and others have been in their own industries.

Analytics and the ability to efficiently and effectively exploit the big data technology stack, advanced statistical modeling, and predictive analytics in support of real-time decision making across business channels and operations will distinguish those companies that flourish in uncertain markets from those that misstep. To avoid missteps, it also helps to have the right technology partners to help get you there.

Next Steps with Dell Technologies and AMD

With the power of data analytics and AI, your company can drive innovations and improve operations. And with trusted technology partners like Dell Technologies and AMD on your side, your AI and analytics-driven journey can be worry-free. With an unrivaled portfolio of modern infrastructure and AI-optimized solutions, Dell Technologies can help you capitalize on the latest technologies — saving you time and money, while reducing risk.

Dell EMC Ready Solutions, powered by the latest AMD, AMD EPYC™ and AMD Instinct Mi100 processors, are engineered with the right balance of performance and low TCO for AI, Analytic and HPC workloads. Built for processing massive amounts of data, the platform features the highest-performance architecture, generous memory and storage options, and industry-leading security.

What's more, Dell Technologies' worldwide HPC and AI Centers of Excellence make it easier to collaborate with some of the brightest minds in AI, HPC, and data analytics. The expert network of resources can help you fine-tune your solutions, test new technologies, and share best practices for optimized results.

Take the next step, today.

Data is everywhere, and it holds the key to better understanding user transactions, customer behavior, machine behavior, security threats, fraudulent activity and more.

Visit our Data Analytics Solutions website or contact your Dell Technologies or authorized partner for more details on how to better leverage your data today.

HPC & AI Innovation Lab

The Dell Technologies HPC & AI Innovation Lab in Austin, Texas, is the flagship innovation center. Housed in a 13,000 square foot data center, it gives you access to thousands of Dell EMC servers, three powerful HPC clusters, and sophisticated storage and network systems. It's staffed by a dedicated group of computer scientists, engineers and subject matter experts who actively partner and collaborate with customers and other members of the HPC community. The team engineers HPC and AI solutions, tests new and emerging technologies, and shares expertise including performance results and best practices.

HPC & AI Centers of Excellence

As data analytics, HPC and AI converge and the technology evolves, Dell Technologies worldwide HPC & AI Centers of Excellence provide thought leadership, test new technologies and share best practices. They maintain local industry partnerships and have direct access to Dell and other technology creators to incorporate your feedback and needs into their roadmaps. Through collaboration, Dell Technologies HPC & AI Centers of Excellence provide a network of resources based on the wide ranging know how and experience in the community.