

Case Study

Hadoop in Banking: The Game Changer

What is Hadoop?

Apache Hadoop is an open-source software framework written in Java for distributed storage and processing of very large datasets on multiple clusters. Developed by Doug Cutting and Mike Cafarella in 2005, the core of Apache Hadoop consists of a Hadoop Distributed File system for storage and MapReduce for processing data. The basic philosophy of Hadoop is to reduce dependence on expensive legacy system hardware to enable distributed parallel processing of very large amounts of data across inexpensive, standard, commodity servers to process and store data without any volume limitations. Hadoop makes the process of storing and managing data economical and reliable.

What are the key features of Hadoop?

- **Reliable:** Fail-Safe technology that prevents loss of data even in an event of hardware failure.
- **Powerful:** Unique storage method based on distributed file system resulting in the faster data processing.
- **Scalable:** stores and distributes datasets to operate in parallel, allowing businesses to run applications on thousands of nodes.
- **Cost-effective:** Runs on commodity machines & network
- **Simple and flexible APIs:** Enables a large ecosystem of solution providers such as log processing, recommendation systems, data warehousing, fraud detection, etc.

How is it revolutionizing the Finance Industry?

Big Data is no longer just a buzzword for the banking and financial industry, for it addresses various issues with the 5Vs; Volume, Velocity, Variety and Veracity. Wondering what the 5th V means? As Bernard Marr, a leading business and Data Expert explains, it is the most important V of big data, for it defines how the other attributes work; Value. Value refers to an organization's ability to turn data into something worth more. It is this Value addition that makes Big Data and Hadoop not just a new trend but a breakthrough for the finance industry.

Banks like BNY Mellon, Morgan Stanley, Bank of America, Credit Suisse, PNC, etc. are already working on strategies around Big Data in Banking, and other banks are rapidly catching up.

How is it used in banking and finance?



The reason for Hadoop's success in the banking and finance domain is its ability to address various issues faced by the financial industry at minimal cost and time. Despite the various benefits of Hadoop, applying it to a particular problem needs due diligence. Some of the scenarios in which it is used are:

Fraud detection:

Fraud, financial crimes and data breaches are some of the most costly challenges in the industry. Hadoop analytics help financial organizations detect, prevent and eliminate internal and external fraud, as well as reduce the associated costs. Analyzing points of sale, authorizations and transactions, and other data points help banks identify and mitigate fraud. For example, big data technology can alert the bank that a credit or debit card has been lost or stolen by picking up on unusual behaviour patterns. This then gives the bank time to put a temporary hold on the card while contacting its account owner.

Risk management:

Every financial firm needs to assess risk accurately, and big data solutions enable them to do so by effectively evaluating credit exposures. Banks analyze transactional data to determine risk and exposure based on simulated market behaviour, scoring customers and potential clients. Hadoop solutions allow for a complete and accurate view of risk and impact, enabling firms to make the best, most informed decisions.

Contact centre efficiency optimization:

Ensuring customers are satisfied is of utmost importance when it comes to finances, and big data can help resolve problems quickly by allowing banks to anticipate customer needs ahead of time. Analyzing data within the contact centre provides agents with timely and concise insight that satisfies customers quickly and efficiently, ensuring cost-effectiveness and even improving cross-sales success rates.

Customer segmentation for optimized offers:

Big data provides a way to understand customers' needs at a granular level so that banks and financial organizations can deliver targeted offers more effectively. In turn, these more personalized offers result in higher acceptance rates, increased customer satisfaction, higher profitability and greater retention. Detailed information about customers derived from social media and transactions can be utilized to reduce customer acquisition costs as well as turnover.

Customer churn analysis:

Everybody knows that it's cheaper to keep a customer than it is to go out and find a new one. Big data and Hadoop technologies can help financial firms keep retain more of their customers by analyzing behaviour and identifying patterns that lead to customer abandonment. When are customers most likely to leave for the competition, and why? What causes customer dissatisfaction? Where did the firm fail? This information for determining how to avoid customer abandonment is priceless. It's imperative for financial firms to learn the right steps to implement in order to meet customer needs and save their most profitable customers.

Sentiment analysis:

Hadoop and advanced analytics tools help analyze social media in order to monitor user sentiment of a firm, brand or product. If a bank is running a campaign, big data tools can monitor social media by name and report on it by hashtag, campaign name or platform. Analytics on the fine-grained details are insightful, and the bank could then make decisions more accurately based on these insights in terms of timing, targeting and demographics.

Customer experience analytics:

As consumer-facing enterprises, financial institutions need to take advantage of the customer data that resides in all of the silos across various lines of business. These include portfolio management, customer relationship management, loan systems, contact centre, etc. Big data can provide better insight and understanding, allowing firms to match offers to a customer or prospect's needs. This then helps the firm to optimize and improve profitable and long-term customer relationships.

Data Storage and Security:

Protection, easy storage and access to financial data are the optimal needs of banks and finance firms. While Hadoop Distributed File System (HDFS) provides scalable and reliable data storage designed to span large clusters of commodity servers, MapReduce processes each node in parallel, transferring only the package code for that node. This means information is stored in more than one cluster but with additional safety to provide a better and safer data storage option.

Does Hadoop have limitations too?

Although Hadoop has been embraced by several banking organizations and it forms the backbone of several applications running Big Data technology, there are also several reasons Hadoop may not always be the best solution. Some of them are:

Big Data understanding:

Hadoop is normally implemented when Big Data is to be implemented. But before using it, one must ask the right questions and ponder upon whether it is the right solution. Any organization that has a huge inflow of data from various sources and which is facing issues to store and effectively use the existing data can use Hadoop and Big Data solutions for their enterprise.

It is not a solution, but a tool:

Hadoop is not the complete solution. Although fraud detection and risk management leverage the strengths of Hadoop, Hadoop by itself does not solve these issues. Programmers need to write codes with an understanding of the problem so that they utilize Hadoop's strong points to solve the business problem. e.g. Big Data does not help in picking up unusual patterns. Big data merely allows large data to be processed concurrently.

Not a unique service:

Hadoop allows analysis, but there are many products that allow you to do data analysis. So, though Hadoop can be used for the purpose of analysis, implementing the framework only to address analytical issues will not be a smart idea. Hadoop is beneficial only if one finds more than one scenario where its USPs can be used properly.

Other vulnerabilities:

Like any other technology, Hadoop is not foolproof or foolproof for that matter. Data is at risk as encryption is missing in the Hadoop system at storage and network levels. Also, since Hadoop makes various duplicates to store data so that it can be retrieved in case of failure, it is vulnerable to data breaches like Java, the language in which the framework of Hadoop is written.

Here's an analogy: Consider a dinner knife. You can use it very well to butter your toast, to cut a piece of potato, to wedge open a shut-in, or use it as a screwdriver for wide-notched screws; but it is useless if you want to drink soup or if you want to make a phone call.

Hadoop is like a knife. The programmers use it to do things effectively where it is applicable. Hadoop does not do fraud detection or risk management, actually, it does not do any business logic by itself; it just manages the storage and retrieval of data in a distributed way.

Conclusion:

The bottom line is that all enterprises, especially financial firms, need to use big data and Hadoop technologies to their fullest potential now, particularly with the overwhelming amount of data and transactions amassed on a daily basis. In order to remain competitive and maintain current customers while attracting new ones, financial firms should start planning to utilize big data technologies today or risk losing more customers to competitors utilizing these tools. That doesn't necessarily mean in every way possible – it just means in the best way possible for each organization.

Big data and Hadoop technologies are powerful and help financial organizations stay ahead in the market. Set them in motion and watch them deliver results.