

Improving operational efficiencies using Big Data for Financial Services

T.AshaLatha¹, Naga Lakshmi.N²

¹ Asst.Professor, Department of Information Technology, Anurag Group of Institutions
Hyderabad, T.S-500 097, India

²Asst.Professor, Department of Information Technology, Anurag Group of Institutions
Hyderabad, T.S-500 097, India

Abstract: Financial services organizations around the world are experiencing drastic change. Financial services firms are turning to big data technologies and Hadoop to reduce risk, analyze fraud patterns. Ability of Big Data platforms to process and analyze unstructured data efficiently unlike current platforms that are limited to handling structured data. This flexibility makes it possible to gain insights from different variety of data. Big data can be analyzed for insights that lead to better decisions and strategic business moves. For most organizations, big data is the reality of doing business. It's the proliferation of structured and unstructured data that floods your organization on a daily basis – and if managed well, it can deliver powerful insights. Big data analytics is the process of examining large data sets containing a variety of data types -- i.e., big data -- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits.

Keywords: Big data analytics, Fraud Detection, Hadoop Clusters, Tools

I. Introduction

Big data analytics

The primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modelers and other analytics professionals to analyze large volumes of transaction data, as well as other forms of data that may be untapped by conventional business intelligence (BI) programs. That could include Web server logs and Internet clickstream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call detail records and machine data captured by sensors connected to the Internet of Things[1]. Some people exclusively associate big data with semi-structured and unstructured data of that sort, but consulting firms like Gartner Inc. and Forrester Research Inc. also consider transactions and other structured data to be valid components of big data analytics applications.

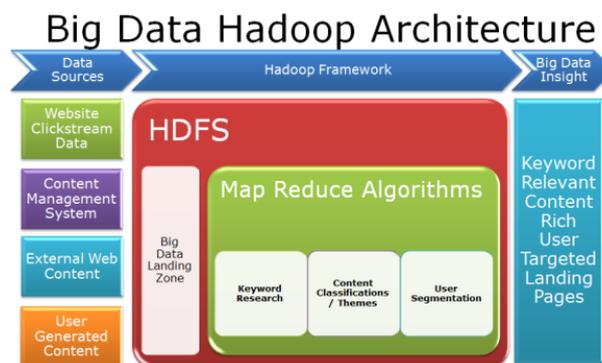


Fig-1

Fraud Detection: Flagging anomalous activities in real time can help prevent potential security attacks or fraud. The MapR Converged Data Platform gives banks the ability to build usage models of “normal” behavior from histories of consumer behavior, analyze incoming transactions against individual and aggregate purchasing histories and take appropriate action if the activity falls outside the confidence level of normal behaviour[2]. As more data is ingested, more precise models can be built so the system can more accurately separate the atypical but legitimate behavior from the suspicious activities.

Risk Aggregation: Big data techniques can be used to gather and process risk data in order to 1) satisfy risk reporting requirements, 2) measure financial performance against risk tolerance, and 3) slice and dice financial reports. The MapR Converged Data Platform can benefit risk managers as they can perform on-demand historical analysis of risk data as well as receive real-time alerts when limits are breached.

Counterparty Risk Analytics: Whenever a firm engages in a business transaction with another party, the risk of doing business with that party must be priced into the terms of the deal[3]. Since calculating counterparty risk requires more than computing a formula, firms typically run long and complex “Monte Carlo simulations” to get a complete picture of risk exposure at many points in time in the future. These simulations require huge volumes of data, massive parallel compute power, and system reliability to ensure firms can continue with business operations with no downtime. The MapR Converged Data Platform provides the performance, scalability, reliability, and the easy access and delivery of data to drive the key components of a counterparty risk analytics system.

Hadoop Clusters A Hadoop cluster is a special type of cluster that is specifically designed for storing and analyzing huge amounts of unstructured data. A Hadoop cluster is essentially a computational cluster that distributes the data analysis workload across multiple cluster nodes that work to process the data in parallel. Hadoop clusters are known for boosting the speed of data analysis applications[4]. The primary benefit to using Hadoop clusters is that they are ideally suited to analyzing big data. They also are highly scalable: If a cluster's processing power is overwhelmed by growing volumes of data, additional cluster nodes can be added to increase throughput. Hadoop clusters also are highly resistant to failure because each piece of data is copied onto other cluster nodes, which ensures that the data is not lost if one node fails.

Big Data improving operational efficiencies:

- Improvement of business process efficiencies by centralizing the processing of data
- Use of faster and more efficient architectures to process data
- Availability of more data for analysis
- Improvements in time-to-market capabilities
- Use of the power of Hadoop to deal with semi-structured information like proprietary XML Blobs eliminating the need for data to be decomposed into 3rd normal form. Use of technologies like HBase since it has a loosely typed schema and decomposing of hierarchical data into column families.
- Aggregation and collation information from multiple lines of business (LOBs) and multiple formats (click stream, app logs, call center log, enterprise data etc.) which can be used to improve prediction-someone defaulting on credit card payments can potentially default on auto or home loans also. Banks can use this data to understand the customer’s needs better (using click stream / call center logs) which would help in deciding the next course of action to be taken
- Emergence of two patterns in terms of usage of Big Data technologies. Banks have built significant assets in terms of SQL queries for analysis over a period of time. NoSQL/columnar databases like Exadata, Netezza and Vertica support ANSI SQL on these appliance and software oriented Big Data platforms[5]. Adoption of such platforms is increasing in cases where banks want to reuse their SQL assets as they run exponentially faster on such platforms.
- The other pattern of usage has to do with the non-consumption end of the data management lifecycle where data has to be provisioned from various source systems and then processed in the Big Data framework layer[6]. In such cases we see the adoption of Hadoop and its ecosystem of products.

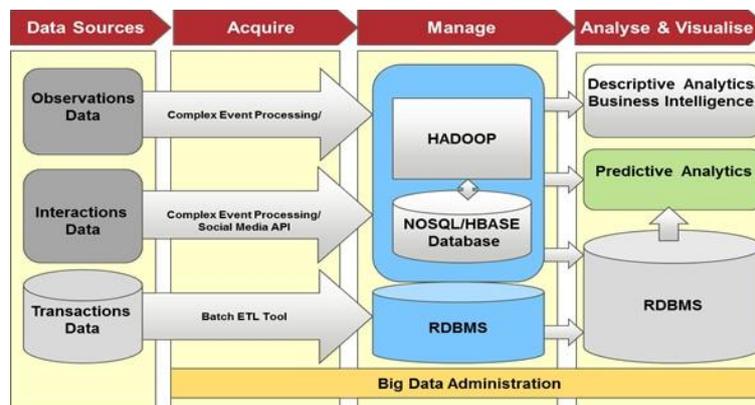


Fig-2

Tools: Big data can be analyzed with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics, data mining, text analytics and statistical analysis. Mainstream BI software and data visualization tools can also play a role in the analysis process. But the semi-structured and unstructured data may not fit well in traditional data warehouses based on relational databases. Furthermore, data warehouses may not be able to handle the processing demands posed by sets of big data that need to be updated frequently or even continually -- for example, real-time data on the performance of mobile applications or of oil and gas pipelines. As a result, many organizations looking to collect, process and analyze big data have turned to a newer class of technologies that includes Hadoop and related tools such as YARN, MapReduce, Spark, Hive and Pig as well as NoSQL databases. Those technologies form the core of an open source software framework that supports the processing of large and diverse data sets across clustered systems.

References

- [1]. Magoulas, Roger; Lorica, Ben (February 2009). "Introduction to Big Data". Release 2.0. Sebastopol CA: O'Reilly Media (11).
- [2]. Hu, Han; Wen, Yonggang; Chua, Tat-Seng; Li, Xuelong (2014). "Towards scalable systems for big data analytics: a technology tutorial". *IEEE Access*. **2**: 652–687. doi:10.1109/ACCESS.2014.2332453.
- [3]. Wu, D., Liu, X., Hebert, S., Gentzsch, W., Terpenney, J. (2015). Performance Evaluation of Cloud-Based High Performance Computing for Finite Element Analysis. Proceedings of the ASME 2015 International Design Engineering Technical Conference & Computers and Information in Engineering Conference (IDETC/CIE2015), Boston, Massachusetts, U.S.
- [4]. Wu, D.; Rosen, D.W.; Wang, L.; Schaefer, D. (2015). "Cloud-Based Design and Manufacturing: A New Paradigm in Digital Manufacturing and Design Innovation". *Computer-Aided Design*. **59** (1): 1–14. doi:10.1016/j.cad.2014.07.006.
- [5]. Lee, Jay; Bagheri, Behrad; Kao, Hung-An (2014). "Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics". *IEEE Int. Conference on Industrial Informatics (INDIN) 2014*.
- [6]. Lee, Jay; Lapira, Edzel; Bagheri, Behrad; Kao, Hung-an. "Recent advances and trends in predictive manufacturing systems in big data environment". *Manufacturing Letters*. **1** (1): 38–41. doi:10.1016/j.mfglet.2013.09.005.