Implementation of Big Data in Commercial Banks

Adeniyi Akanni, Ph.D., CISA, CRISC, ITIL
Worked as an experienced systems auditor and an information security expert in commercial banks in Nigeria for more than two decades, holding various positions as head of e-fraud and investigation and head of reconciliation and compliance. He currently works with the First Bank of Nigeria Ltd.

Big data can be described as a huge volume of data that cannot be treated by traditional data-handling techniques.[1] Considering the enormity of data generated in various forms at various times via various devices, it is clear that such data would not only be unstructured, but also complex. Without proper collation, coordination and harnessing of those data, meaningful decisions may not be reached by relevant users, such as chief information technology officers, systems auditors and chief executive officers (CEOs).

Proper implementation of big data can be an indicator of effective usage of big data because data continue to grow exponentially.[2] Big data is big because of a high level of volume, velocity and variety. This high level is due to the way data are generated and continuously increasing in quantity (volume) at a very fast rate (velocity) and in various forms (variety).

The cost associated with storing petabytes of data is not the major problem for organizations such as commercial banks, because cloud service providers (CSPs) can offer such services at reasonable prices. The big challenge lies in the form in which the data are generated, which does not follow a specific pattern (relating to the variety of data generation), and the rate at which the large quantity of data is generated (velocity).[3] Thus, a careful selection of strategy is necessary so that such data can facilitate informed decision making that would, in turn, affect the security of the data and positively leverage information obtained from big data for a competitive edge.[4] This article describes a six-stage cycle of implementing big data in commercial banks, points out the major challenges in implementation and provides a suggested solution. It also assists CEOs with properly analyzing their data for optimal marketing drives.

Implementation of Big Data at the Bank's End

Implementation of big data involves a well-planned strategy for organizations to get the best out of it and make informed decisions that will guide their marketing drives. It can be seen as a six-stage cycle that involves definition, identification, recognition, analysis, ploughing back past experiences and training. These are critical stages to consider in big data implementation because each of them will help implementers to be focused on the expected result of the stage and the ultimate goal of the organization in implementing big data. This is illustrated by the acronym DIRAPT (**figure 1**). It should be seen as a cycle that every organization needs to repeat.

Definition of Scope

A major perception is that big data is seen as boundless. It is true that data generated by banks are enormous, ranging from daily interbranch transactions to messages sent or received via various media. The amount of data is determined by the number of branches, staff members and customers.

Sometimes, commercial banks, based on marketing focus and strategy, may decide to streamline the market in which they play. For instance, one commercial bank may be public-sector-focused while another may choose to focus on retail banking. Scope also invariably determines how large their customer base will be. For instance, banks whose strategy focuses on the retail sector will likely have more customers than those dealing with public or corporate sectors. Besides transactions emanating from various branches, there are interbank activities involving the movement of funds

Figure 1—DIRAPT Cycle
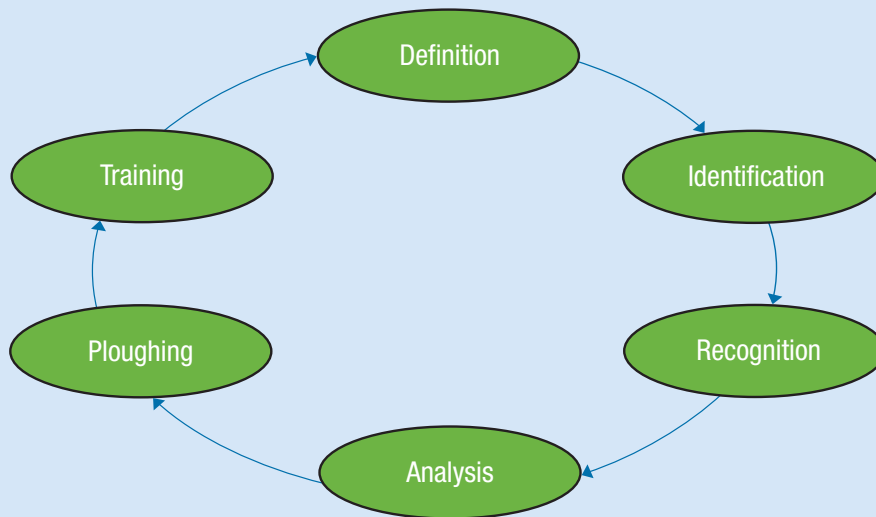
through alternative channels, such as automated teller machines (ATMs) and mobile banking, where large chunks of data are exchanged. It will not make any sense to want to treat all data from all fronts the same way and at the same time. It is necessary for banks to define the scope of big data implementation to be covered in order to get meaningful information from the data.

Identification of Skill Set

After developing a successful definition of boundaries in which to work, it is necessary to identify human resources needed with the required skill set. Careful selection of manpower with the requisite skills is very important before a successful big data implementation. Banks should note that this should not be seen as residing in only one department of the bank. Experienced staff should be picked from operations, marketing, control and other departments to contribute their input for successful implementation. A rich blend of skilled people will go a long way to determine the success of an implementation.

Recognition of Data Sources

Effective data tracking and measurement stem from identified data sources. It has been said that if it

cannot be measured, then it cannot be controlled. Each data source must be listed, although not all of them can be handled at once. Then, based on the defined scope, data from the identified sources can be prepared for the next stage.

> " EFFECTIVE DATA TRACKING AND MEASUREMENT STEM FROM IDENTIFIED DATA SOURCES. IT HAS BEEN SAID THAT IF IT CANNOT BE MEASURED, THEN IT CANNOT BE CONTROLLED. "

Analysis of Output

Analysis is the stage where data within the scope are reviewed for relevant information for management use. Both structured and unstructured

Training and Retraining

Training is a continuum. There should be regular training for bank workers involved in implementation before, during and after each cycle of implementation. Lessons learned at every stage should be well coordinated and recorded for reference purposes. Training should be encouraged at every stage of the DIRAPT cycle as well as at the end of each cycle.

Challenge With DIRAPT Cycle Approach

A major challenge with big data implementation is cost, in terms of human resources and infrastructure. Hiring people with relevant skills is always a Herculean task because knowledgeable personnel who can handle big data are scarce. On the technology side, the cost of getting appropriate tools to handle big data is also high. However, the DIRAPT cycle has brought in cloud solutions to help enterprises reduce the huge cost with the emergence of CSPs offering their services at reasonable prices.

Practical Approach to Big Data Implementation by CSPs

Commercial banks run their day-to-day activities using various software applications. Different segments of commercial banks (e.g., local operations, foreign operations, credit risk management, internal audit, internal control and information technology) run their applications in silos. Software contains useful information. Bringing together the data generated by all departments for the organization is always a challenge.

Recent technological advancements have helped bring big data into the cloud. Several CSPs now implement big data by using Software as a Service (SaaS). At the CSP end, activities are broken down into four major areas: the enablement, migration, integration and testing (EMIT) approach.[5] Most CSPs use the EMIT approach to handle big data.

Enablement
CSPs tend to enable business applications used by banks to operate on their own environment rather than the banks', which is basically on SaaS. At this

data are involved. While the former is not much of an issue, the latter may require specialized analytics tools such as Hadoop, NoSQL, Jaspersoft, Pentaho, Karmasphere, Talend Studio and Skytree for analysis. A good analysis of data helps management make informed decisions to move the organization forward.

> " A MAJOR CHALLENGE WITH BIG DATA IMPLEMENTATION IS COST, IN TERMS OF HUMAN RESOURCES AND INFRASTRUCTURE. "

Ploughing Back Experience

Experience is a very important aspect of big data utilization. To begin, efforts should be made to get experienced personnel when implementing big data. Over time, experience gained can be expanded and reused. No two projects will be exactly the same, but experience gathered from a previous project can always be considered in subsequent projects. Therefore, it makes sense for experienced staff to be used in subsequent iterations for optimal results.

stage, processes and workflows are configured and allowed to run on the CSP environment as if they were offered to run the business applications for their clients—the commercial banks.

Migration
CSPs do a customized transfer of data to appropriately and cost-effectively move data to the cloud. This stage takes the place of the traditional data center and its infrastructures. By doing this, the cloud solution could help reduce costs.

Integration
The integration phase helps to incorporate both applications and data of the banks with the CSP's interface. Thereafter, a development area is prepared for banks to run. Banks first operate in the development area to properly assess the integration. At the bank's end, a user accesses departmental applications, through the interface provided, once a connection is established.

Testing
All business applications are tested before go-live. This includes interface testing, unit testing, stress testing, functional testing and performance testing of the various applications running on the cloud.

Conclusion

Big data is a concept that has received much publicity recently, possibly because of the intricacies involved. Efficient use of data in bank settings is very important to dictate how far and fast a commercial bank can go in the next few years.[6] Thus, there is a need for banks that hope to thrive to have a proper understanding of their data through a carefully selected big data implementation strategy.

The DIRAPT cycle spells out ways of implementing big data in commercial banks to help them enjoy the derivable benefits of big data, which include, but are not limited to, data security and competitive advantage. On one hand, proper implementation of big data will help banks to discover security risk regarding data exposure to fraudulent manipulations and then devise appropriate measures to mitigate them. Second, banks stand the chances of designing appropriate products for the appropriate environment with a view that can help them outwit their competitors.

Endnotes

1 Mukherjee, S.; R. Shaw; "Big Data—Concepts, Applications, Challenges and Future Scope," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, iss. 2, February 2016
2 Ramlukan, R.; "How Big Data and Analytics Can Transform the Audit," *Reporting*, iss. 9, April 2015
3 Watson, H.; "Big Data Analytics: Concepts, Technologies and Applications," *Communications of the Association for Information Systems*, vol. 34, 2014
4 Charles, V.; T. Gherman; "Achieving Competitive Advantage Through Big Data," *Middle-East Journal of Scientific Research*, vol. 16, iss. 8, 2013, p. 1069–1074
5 Orion, "Grow With a Connected Business," *www.orioninc.com/cloud/*
6 Ball, S.; "Gartner's 2017 BI Magic Quadrant: What Differentiates the Leaders?," *Better Buys*, 5 June 2017, *https://www.betterbuys.com/bi/gartners-2017-bi-magic-quadrant/*