Big-data in Mobile Networks

O. Théophile Aballo, Roland Déguénonvo and Antoine Vianou

***Abstract*—** With the growing number of mobile phone users and opportunities to use the Internet from SIM (Subscriber Identity Module) cards, mobile network operators are facing massive user data management. A fast processing application is needed to solve this problem. The implementation of Bigdata was presented as an effective solution, it consists in reducing the risks of unnecessary storage of data and facilitating operator decision-making thanks to predictive analysis and a much more personalized and contextualized customer experience. In very large petabyte and zetabyte networks, Bigdata can manage data that no traditional database management or information management tool can really work with. The purpose of this article is to study the Bigdata tools that help to increase the performance of mobile networks.

***Keywords*—** Mobile networks, massive data management, storage, analysis

I. Introduction

The development of the Internet nowadays engenders a massive use of mobile data that is generated on the systems. However, for companies, there is a problem of data storage on the one hand and a better determination of current trends through the analysis of these stored data [1].. For a mobile operator of five hundred thousand subscribers in 2006, the number of subscribers has increased by five million today. Much more advanced equipment is needed for data storage and new methods for effective treatment [10].. Reports that require a long period of preparation and analysis are required today for very short deadlines. The Bigdata is the miracle solution to fix its problems.

II. Material And Methods

Bigdata technologies could help mobile operators to efficiently analyze data faster than traditional databases. With the Hadoop tool, operators would be able to analyze not only the structured data of their systems, but also the unstructured data generated by their subscribers.

The method consists of collecting and managing data from a mobile operator. It is done in four steps:

A. Number Recycling

Inactive numbers for a period of at least three months are recycled in order to be sold again on the market.

O.T.Aballo, Polytechnic School / University of Abomey-Calavi, Benin.
R. Deguenonvo, Doctoral School of engineering / University of Abomey-Calavi, Benin
A.Vianou, Doctoral School of engineering / University of Abomey-Calavi, Benin

B. Implementation of new numbers

Recycled numbers are installed on different nodes. It is a set of parameters of the authentication center, billing, a service delivery platform and a geographical location recorder subscribers.

C. Registration of the subscriber on the network

A number configured on the network cannot be used directly. It is installed with an OBO (Barring of all outgoing calls) barrier. Only the registration made on the number allows to raise this barrier.

D. The recording of data generated by Network traffic

Thanks to the telecommunication servers, as the traffic is generated, logs are inserted into tables of different databases by means of ETL (Extract Transform Load) .These are the different tables that will be later operated for billing, calculating the business income following each service offered, handle the various subscriber complaints, provide indicators for marketing.

III. Resultats

Here we will try to use Bigdata technologies with the pseudo-distributed mode of the cluster presented above. The different cases of uses studied are the following:

- Use case 1: Qualitative study of the 6 million subscribers of a mobile operator.

We have created test datasets based on the Java Random method. The fields generated here are surname, first names, nationality, age, sex, MSISDN, area, profession. We made the restitution of the data via Hive.

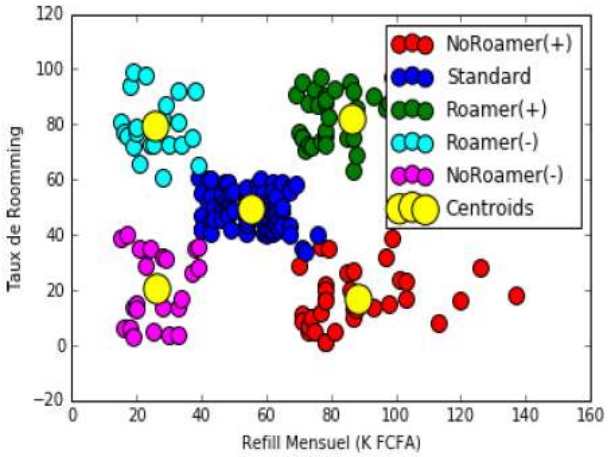

Fig. 1 Segmentation of roaming subscribers.

https://doi.org/10.17758/EIRAI4.F0918115

-Use case 2: Real-time study of subscribers views on social networks.

In this part, we collected data on twitter and tested thanks to the platform developer. Here is a dictionary of key words extracted from the platform:

OPERA Zik - OPERA GO PACK - OPERA WABA - OPERA PROPRE - OPERA Me2U -OPERA Bip Me - OPERA Save Number - OPERA Pay My Call - OPERA Karaoke - OPERA Mobile Money

IV. Discussion

Use case 1: After analyzing the data by {SELECT age, count (*) FROM axon_base GROUP BY age}, fig 1 shows us that our task was executed successfully. In this case three jobs were solicited, six mapper turned in 2 minutes 53 seconds 940 milliseconds.

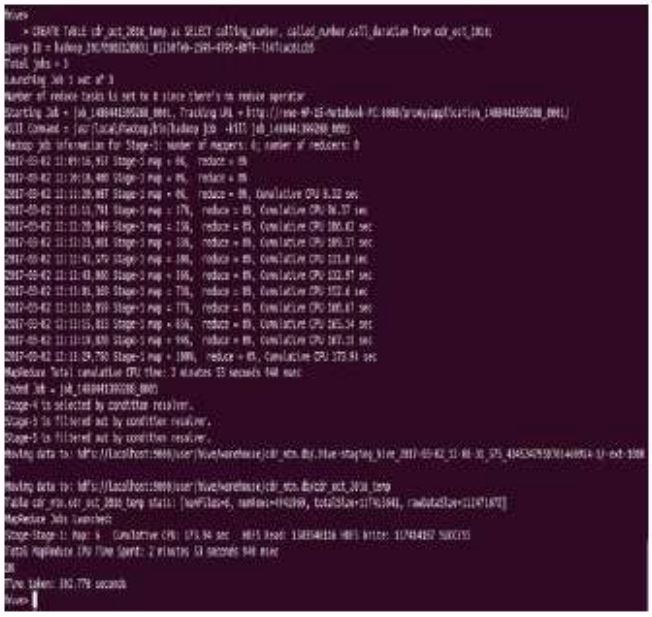

Fig. 2 Running a Job

A much better performance than the oracle database that could even fail if the connection to the storage server was not stable or was heavily solicited. Reading the HDFS file, running the mapper, the reducer and generating the output file took 302.778 seconds. We would spend about 10 minutes exporting in .txt or .csv a large file with an Oracle database architecture. Fig.3 et Fig.4 present an execution report of the previous job.

Fig. 3 Map Running

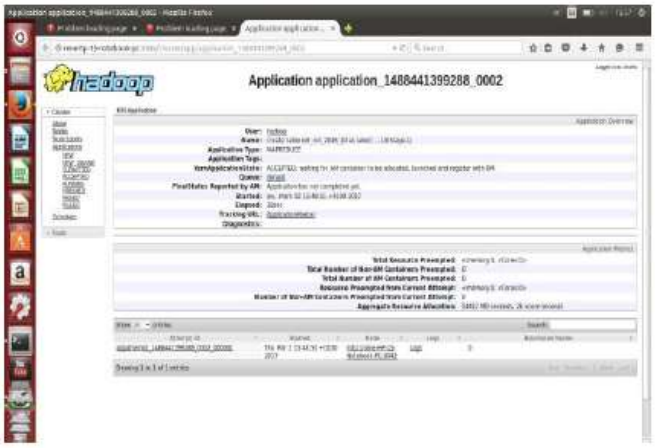

Fig. 4 Task Execution Report

Case of Use 2: it appears that the operator would benefit from getting closer to these people having a very high visibility on this social network in order to have their comments and opinion on the products mentioned in the dictionary of keywords indicated. They are true delegates for their subscribers and they would be a precious help in order to always preserve the friendship and the complicity which remains for more than 10 years with the subscribers of the operator.

V. Conclusion

The quantitative explosion of digital data forces new ways of seeing and analyzing the world. New orders of magnitude concern capturing, storing, searching, sharing, analyzing and visualizing data. The prospects of big data processing are enormous and, for the most part, still unsuspected. There is often talk of new opportunities for media-based information exploration, knowledge and evaluation, trend and prospective analysis and risk management, and religious, cultural, political, and genomics or meta genomics, for medicine, meteorology and adaptation to climate change, complex energy network management, ecology. The multiplicity of these applications already leaves appearing a true economic ecosystem implying, the biggest players of the sector of the information

technologies. Bigdata remains today at the heart of technological challenges. Better management of the volume, speed and variety of data creates significant added value..

REFERENCES

[1] Xavier Lagrange, Philippe Godlewski, Sami Tabbane, « Réseaux GSM », Editions Hermès Science (5ème édition), Paris 2000, ISBN 2-7462-0153-4
[2] J.R. Owens, John Lentz, and Brian Femiano, « Hadoop par la Pratique, » PEAR-SON 2014.
[3] Robert Layton, « Learning Data Mining with Python » Date de Publication originale 29 Juillet 2015
[4] Donald Miner, Adam Shook, MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop, 2012
[5] Chuck Lam, “Hadoop in Action”, Date de publication originale 2011.
[6] Kenneth Cukier, “Learning with. Big Data: The Future of Education”, 2014
[7] Rick Smolan, Jennifer Erwett, “the Human Face of Big Data”, 2012
[8] Ritu Aora, “Conquering Big Data with High Performance Computing”, 2016
[9] Bhuvan Unhelkar, “Bigdata Strategies for Agile Business”, 2017
[10] Krish Krishnana, “Data Warehousing in the Age of Big Data,” 2013