Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 508–512

Procedia Computer Science

www.elsevier.com/locate/procedia

The 2nd International Workshop on Data Mining in IoT Systems (DaMIS 2017)

Analysis of a data-flow in a financial IoT system

Salvatore Cuomo*[a], Vittorio Di Somma[a], Federica Sica[a]

[a] University of Naples Federico II -via Cintia, Napoli 80126,Italy

Abstract

Data retrieving, analysis e management are usually known as complex task in financial contexts. In an Internet of Things (IoT) system data-flow processes represent the knowledge base used in mathematical models for credits and financial products. Several sources such as distributed database systems, portals and local information are generally used as input of inferring models. In this paper we describe an overview of software tools, methodologies and strategies in real data-flow system.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Internet of Things; Financial system; data-flow management.

1. Introduction

In many applicative scenarios processes make use of huge amounts of interrelated data: this explains the necessity of techniques for their classification and managing. Iot frameworks are very suitable to this kind of contexts for several reasons, in particular: i) the diffusion of sophisticated tools (smart phones, tablets and smart watches); ii) the possibility of real time data; ii) efficient *communication models* among devices. As it concerns the point iii), we list some common models: a) *Device-to-Device Communications*, i.e. two or more devices that directly connect and communicate between one another; b) *Device-to-Cloud Communications*, where the IoT device connects directly to an Internet cloud service like an application service provider to exchange data and control message traffic; c) *Device-to-Gateway Model*, where there is an application software operating on a local gateway device, which acts as an intermediary between the device and the cloud service and provides security and other functionality such as data or protocol translation. One of the main opened problems related to IoT systems is a *derivative effect*, i.e. many financial transactions are based on information from intangible sources and only indirectly form real objects (for this reason, scientists research new methods for the improving of IoT sensors and the analysis of the their data). For an introduction to IoT the reader can see [1]. IoT finds several applications in the cultural heritage (see [2],[3],[4],[5],[6],[7],[8]).
Another sector characterized by very big data flows to analyze and from which inferring information is represented by *financial* one. For example, in banking context, IoT applications are able to improve underwriting processes for several

*Corresponding authors. Salvatore Cuomo, Vittorio di Somma, tel. 0039675624
E-mail address: {salvatore.cuomo, vittorio.disomma}@unina.it

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.313

purposes: a) obtaining more information of goods; ii) monitoring the condition of different assets market; iv) helping traders to choose the best opportunity. In particular, all these characteristics can take advantage in the determining the *no arbitrage price of a European option* (this problem is discussed in[9], [10], [11]); for IoT financial description we refer to[12].

2. Financial Data Flow System

In Finance, data can be involved in a complex and long process which allows the financial institution to properly treat and make advantage of them. Data losses, misinterpretation and optimization are one of the main issues financial (and non) institutions must face.

We can wrap up the data-flow in three main phases:

- **Data retrieving**: the financial institution retrieves data from more than one source, often causing operational risk.
- **Data analysis and management**: data are extracted from the database to be used for analysis purposes.
- **Data reporting**: data are transferred again to the main database and/or externally reported.

This general and simplified scheme is represented with more details in figure 1.

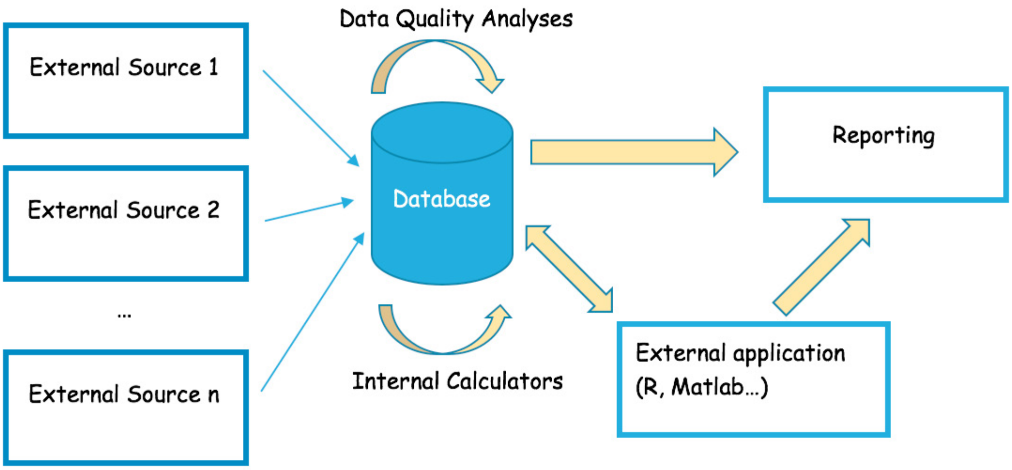

Figure 1. Data-flow.

2.1. *Data Retrieving*

The ideal situation for a financial institution would be gathering all data from only one database. Nowadays, Business Intelligence is taking care of this aspect and is facing the tough challenge of joining old databases in a new, functional one.

Typically, given the huge size of banks and their step by step adjustment to new software and technologies, it is still hard to combine all data, particularly when treating historical ones. This is the reason why also (if not especially) huge financial institutions are still working with more than one database, generating overlapping data which in general do never exactly match.

We will here provide an example of how databases containing data for Risk Management department could be structured. We can imagine a database for credit risk data and a second database for market risk data retrieval.

- **Credit Risk database.**
  This database contains all data for credit risk purposes. Data are gathered, filtered, checked and finally used as input for credit risk models. Eventually, data are reported to other departments or saved in the main database, such that it will be possible to use them again. Given the different goals, various libraries are available in the database.

  1. A first library would contain historical data. It could be a copy of an old database, or just a static collection of tables ready to be read and used.
  2. A second library contains external sources data, that are uploaded and remain static or/and are continuously updated.
  3. Some calculations can be performed directly in the database, so a library will be used to store all the calculators and their results. These data are read to be analyzed.
  4. When analyses are performed, results are stored in a particular library that everyone can access from their local environment.
  5. Since calculations and analyses can be long, it is useful to produce tables in intermediate steps to check the process. Given the large amount of resulting data, it is better to store these intermediate results in a local library, that will be deleted at the end of the process.

  In each library, tables with all data can be selected and filtered or exported. We can assume data are uploaded at different time intervals: new loans can be added daily; calculators perform monthly; etc.

- **Market Risk database.** A Market Risk database will contain more data coming from external sources, often not free. Data are used to monitor traders and bank portfolios, develop and validate market risk models, and price financial instruments with advanced techniques.
  Data will be provided by:
  1. External sources:
     a. Mainly Bloomberg, with a specific add-in that allows to download market risk data directly to the database, minimizing operational risk;
     b. Markit, Superderivatives, etc., id est other market data providers similar to Bloomberg. Excel files can be manually downloaded from their websites.
  2. Internal sources: traders, who upload data regarding every new product they sell or buy. Usually this process is manually done, producing operational risk. Market risk managers check data quality of inserted data.

2.2. *Data Analysis and Management*

Data are processed to be filtered, checked, and eventually used as input for models or analyses. There are different types of software that can be chosen for these tasks. We will present the most common in financial world.

1. **Data Management Tools.**
   To manage databases and perform filters, data quality check or to build basic calculators, financial institutions use SQL Management Studio or/and SAS.

   a. SAS. Useful for data storage and filtering, but commonly preferred because of its predefined models for Statistics and Data Analysis. SQL procedure allows to use SQL syntax; SAS programming language is not easy to fully understand. User-friendly, it can be used even without knowing how to program. Not fast for modelling as other programming languages could be.
   b. SQL Management Studio. Specialized in data storage and filtering, it can easily produce reports. It is free-source and very fast. The syntax is clear and easy to learn.

2. **Data Analysis and Modelling Tools.**
   Various programming languages can be chosen for modelling, depending on the specific requirements. Some languages are faster than others, but they take time to efficiently program.

   a. R/RStudio. The most common for data analysis and Statistics. Free-source, can be adjusted to become object-oriented. Usually fast, it can be faster thanks to parallel computing techniques. $C++$ and Fortran programs can be easily ran in R. It works better with *csv* files. Easily linked to SQL.
   b. MATLAB. Not free-source, but more reliable. It offers packages for parallel computing, global optimization, financial modelling. Already fast and with a user-friendly IDE. Not object-oriented. Easily linked to Excel and SQL.
   c. Python. Free-source, it is fast and very precise. Good for heavy computational tasks.
   d. $C++$ and Fortan. Faster than other languages, but they require more time in programming. Object-oriented. Definitely more precise in computation. They are preferred for heavy computational tasks.

Financial institutions usually prefer to use programs/types of software they have to pay for, since this allows them not to take responsibility on the results of the model. Indeed, if a particular package of a free-source software contains a bug, it will be bank's responsibility to justify the mistake.

2.3. Data Reporting

Eventually, data are ready to be reported to an external institution or to another department.

Excel files are the easiest way to send data and tables, but more sophisticated software can be used. Data-flow can be automatized so that the operational risk of copying and pasting in excel disappears. E.g., reports can be produced directly with RStudio, which performs calculations on input data and then returns dataframes or datatables that can be already passed to the report.

Packages such as *Rmarkdown* and *Knitr* allow to use only one function to create a report, just passing the input parameters to the function. The process is fully described in figure 2: an *.Rmd* file with markdown text (similar to latex syntax) and R code chunks is passed to Knitr, that executes the chunks and creates a new markdown file with extension *.md*. Finally, the *.md* file is processed by *pandoc* that creates the final output: a Word file. Other types of output can be selected, such as a pdf or htlm file, opportunely choosing latex or other document converter.

Figure 2. Rstudio: final output creation.

3. Conclusions

In this paper we have described a financial data flow, which can be divided into the following steps: i) **Data retrieving**: the financial institution retrieves data from more than one source, often causing operational risk; ii)**Data analysis and management**: data are extracted from the database to be used for analysis purposes; iii)**Data reporting**: data are transferred again to the main database and/or externally reported. In the following our purpose consists in applying data obtained for inferring information about some financial problem (for example, for the study of the option pricing problem), by improving our algorithm thanks to the use of sophisticated IT tools (as parallel calculus).

References

1. Kevin, A.. That "internet of things" thing, in the real world things matter more than ideas. *RFID Journal* 2009;**22**.
2. Salvatore, C., De Michele, P., Galletti, A., Piccialli, F.. A cultural heritage case study of visitor experiences shared on a social network. In: *10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 3PGCIC 2015, Krakow, Poland, November 4-6, 2015*. ISBN 978-1-4673-8317-2; 2015, p. 539–544. doi:10.1109/3PGCIC.2015.55.
3. Chianese, A., Marulli, F., Moscato, V., Picciall, i.F.. A "smart" multimedia guide for indoor contextual navigation in cultural heritage applications. IEEE Computer Society. ISBN 9781479940431; 2013, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84904581677&doi=10.1109%2fIPIN.2013.6851448&partnerID=40&md5=a2ed7ba365db19484e6bc1a3ddf8332a. doi:10.1109/IPIN.2013.6851448.
4. Cuomo, S., De Michele, P., Posteraro, M.. A biologically inspired model for describing the user behaviors in a cultural heritage environment. ISBN 9781634391450; 2014, p. 292–302. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84914168950&partnerID=40&md5=226c2f691943390a490404c5bc7827cc.
5. Chianese, A., Piccialli, F.. Smach: A framework for smart cultural heritage spaces. ISBN 9781479979783; 2015, p. 477–484. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84928558329&doi=10.1109%2fSITIS.2014.16&partnerID=40&md5=fd30dc70fe90a12c97427c29e3ac198e. doi:10.1109/SITIS.2014.16; cited By 6; Conference of 10th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2014 ; Conference Date: 23 November 2014 Through 27 November 2014; Conference Code:111857.
6. Cuomo, S., De Michele, P., Galletti, A., Picciall, i.F.. A cultural heritage case study of visitor experiences shared on a social network. Institute of Electrical and Electronics Engineers Inc. ISBN 9781467394734; 2016, p. 539–544. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84964545804&doi=10.1109%2f3PGCIC.2015.55&partnerID=40&md5=e7f883939b4aec336b79d3971ed5eda5. doi:10.1109/3PGCIC.2015.55.
7. Chianese, A., Piccialli, F., Riccio, G.. Designing a smart multisensor framework based on beaglebone black board. *Lecture Notes in Electrical Engineering* 2015;**330**:391–397. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84915822524&doi=10.1007%2f978-3-662-45402-2_60&partnerID=40&md5=9a82cf44d19cc5b470240f10ee683342. doi:10.1007/978-3-662-45402-2_60; cited By 12; Conference of 6th FTRA International Conference on Computer Science and its Applications, CSA 2014 ; Conference Date: 17 December 2014 Through 19 December 2014; Conference Code:111689.
8. Cuomo, S., De Michele, P., Piccialli, F., Galletti, A., Jung, J.. Iot-based collaborative reputation system for associating visitors and artworks in a cultural scenario. *Expert Systems with Applications* 2017;**79**:101–111. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85014468154&doi=10.1016%2fj.eswa.2017.02.034&partnerID=40&md5=7368e462f1824e0d6c59caef8f2c26fd. doi:10.1016/j.eswa.2017.02.034; cited By 1.
9. Cuomo, S., De Michele, P., Di Somma, V., Galletti, A.. A stochastic method for financial iot data. *Procedia Computer Science* 2016; **98**:491 – 496. URL: http://www.sciencedirect.com/science/article/pii/S187705091632227X. doi:http://dx.doi.org/10.1016/j.procs.2016.09.082; the 7th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2016)/The 6th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2016)/Affiliated Workshops.
10. Cuomo, S., De Michele, P., Di Somma, V., Ponti, G.. A probabilistic approach for financial iot data. vol. 1774. 2016, p. 73–74. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85013392637&partnerID=40&md5=fa77e1e974e9c24961dae7d45212aa0d.
11. Cuomo, S., Campagna, R., Di Somma, V., Severino, G.. Numerical remarks on the estimation of the option price. 2017, p. 746–749. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019215390&doi=10.1109%2fSITIS.2016.123&partnerID=40&md5=75088fabd717e3d8ef2434d4ddd7e792. doi:10.1109/SITIS.2016.123.
12. Eckenrode, J.. The derivative effect: How financial services can make iot technology pay off. *Deloitte University Press* 2015;.