

## Chapter 9

# A Framework to Evaluate Big Data Fabric Tools

**Ângela Alpoim**

*University of Minho, Portugal*

**João Lopes**

*University of Minho, Portugal*

**Tiago Guimarães**

*University of Minho, Portugal*

**Carlos Filipe Portela**

*University of Minho, Portugal*

**Manuel Filipe Santos**

*University of Minho, Portugal*

### ABSTRACT

*A huge growth in data and information needs has led organizations to search for the most appropriate data integration tools for different types of business. The management of a large dataset requires the exploitation of appropriate resources, new methods, as well as the possession of powerful technologies. That led the surge of numerous ideas, technologies, and tools offered by different suppliers. For this reason, it is important to understand the key factors that determine the need to invest in a big data project and then categorize these technologies to simplify the choice that best fits the context of their problem. The objective of this study is to create a model that will serve as a basis for evaluating the different alternatives and solutions capable of overcoming the major challenges of data integration. Finally, a brief analysis of three major data fabric solutions available on the market is also carried out, including Talend Data Fabric, IBM Infosphere, and Informatica Platform.*

## **1 INTRODUCTION**

Today, it is essential for the success of organisations to “be smart”. It can translate into quick and agile decisions, turning information into knowledge that can help make the best decisions (Zikopoulos and Eaton 2011). Data are generated, analysed and used on an unprecedented scale, and decision-making is being applied to all aspects of society (Srivastava 2013). Never have so many records been generated about what people do, think, feel or desire as they do today. Therefore, people’s daily interactions with widespread systems create traces that capture various aspects of human behaviour, allowing different machine learning algorithms to extract valuable information about users and their actions.

The management and analysis of this huge amount of data, through the analysis of banking transactions, online surveys, access to websites or even with the appearance of connected devices such as smartphones or smartwatches, can be seen, simultaneously, as one of the most significant benefits and challenges of organizations. It is as important to obtain and generate information as being able to process it quickly (Volpato et al. 2014). And this is one of the greatest challenges: organising and modelling the data to facilitate the process of linking, transforming, processing and analysing the data collected in order to make the best decisions promptly (Cassavia et al., 2014). This case requires the exploitation of adequate resources, new methods, as well as the ownership of the appropriate technology (Oussous et al. 2017). The truth is that the process of selecting appropriate integration tools for different types of businesses is crucial, given the growing demand and need from data and information companies. However, it is important to realize two important factors before moving to a major data project and implementing a major data integration solution. According to a framework developed in a previous study (Portela et al. 2016), one of the most important aspects in this process is to frame the existing problem with two important questions:

1. Is it really a big data problem?
2. Is it really necessary to have large data tools to solve the problem in question?

These questions need to be assessed before choosing and investing in a large data solution, because even this investment can be seen as being at risk for many companies.

In this sense, after the correct framing of the project in its real dimension, and if it assumes the structure of a large data project, it is possible to implement the structure of BigDAF by introducing a tool analysis component, namely the Evaluation Model, which was developed to guide the decision and classification of tools taking into account the different requirements, needs and evaluation criteria of the different users, in order to select the best data integration solution, according to the needs of each company.

This research is therefore directed in the context of identifying and analysing some of the solutions existing in the market, adopting an Evaluation Model that can support the assessments developed, with the main objective of recommending the best option according to previously defined criteria, representing the most important decisions to be taken into account within the existing problem. In fact, this paper extends a study performed about Big Data Integration Tools (Alpoim et. all 2019).

This document is structured in seven sections. The first section provides a brief introduction to contextualize this study. The second section describes and characterises the concept of large data, presenting the data tools as well as the BigDAF structure concept. Section three represents the solution for evaluating large data integration tools, describing the main criteria that can be used to make this evaluation. Section four describes the results of this study and an example of the application of the evaluation model

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-framework-to-evaluate-big-data-fabric-tools/267873](http://www.igi-global.com/chapter/a-framework-to-evaluate-big-data-fabric-tools/267873)

## Related Content

---

### Randomizing Efficiency Scores in DEA Using Beta Distribution: An Alternative View of Stochastic DEA and Fuzzy DEA

Parakramaweera Sunil Dharmapala (2014). *International Journal of Business Analytics* (pp. 1-15).  
[www.irma-international.org/article/randomizing-efficiency-scores-in-dea-using-beta-distribution/119494](http://www.irma-international.org/article/randomizing-efficiency-scores-in-dea-using-beta-distribution/119494)

### Price Discounts and Consumer Load-Shifting Behavior in the Smart Grid

Eeyad Al-Ahmadi and Murat Erkoç (2018). *International Journal of Business Analytics* (pp. 33-54).  
[www.irma-international.org/article/price-discounts-and-consumer-load-shifting-behavior-in-the-smart-grid/192167](http://www.irma-international.org/article/price-discounts-and-consumer-load-shifting-behavior-in-the-smart-grid/192167)

### National Intellectual Capital Stocks and Organizational Cultures: A Comparison of Lebanon and Iran

Jamal A. Nazari, Irene M. Herremans, Armond Manassian and Robert G. Isaac (2010). *Strategic Intellectual Capital Management in Multinational Organizations: Sustainability and Successful Implications* (pp. 95-118).  
[www.irma-international.org/chapter/national-intellectual-capital-stocks-organizational/36458](http://www.irma-international.org/chapter/national-intellectual-capital-stocks-organizational/36458)

### Biological and Medical Big Data Mining

George Tzanis (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 246-262).  
[www.irma-international.org/chapter/biological-and-medical-big-data-mining/142621](http://www.irma-international.org/chapter/biological-and-medical-big-data-mining/142621)

### Predictive Analytics and Data Mining: A Framework for Optimizing Decisions with R Tool

Ritu Chauhan and Harleen Kaur (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 359-374).  
[www.irma-international.org/chapter/predictive-analytics-and-data-mining/142628](http://www.irma-international.org/chapter/predictive-analytics-and-data-mining/142628)