Big data based architecture to bringing together graduates and recruiters: case of Moroccan university

Abdemounaime Hamdane, Nadir Belhaj, Halima El Hamdaoui, Karima Aissaoui, Moulhime El Bekkali, Nour El Houda Chaoui
Department of Electrical and Computer Engineering, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Article Info

Article history:
Received Jun 30, 2021
Revised Mar 16, 2022
Accepted Apr 3, 2022

Keywords:
Big data
Company’s database
Data warehouse
Generated data
Graduates database
Unemployment
University

ABSTRACT
Due to the current health crisis caused by COVID-19, a negative impact has occurred on the global economy and more specifically on employability. Many people have lost their jobs or have seen their incomes drop. Nowadays, the search for job offers or potential candidates is done mainly online, where several platforms already exist (LinkedIn, Viadeo or others online recruitment systems). These solutions are particularly difficult to use due to the volume of data to be found and the manual compatibility check. In addition, the surplus of unqualified candidates and unverified resumes is a major concern of online recruiting systems. What we propose in this article is a framework that helps bridge the gap between graduates and recruiters through a big data architecture for university based on a real and certified database of graduates and companies.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION
The question of the integration of young graduates remains an open debate, which requires the involvement and interest of all stakeholders and all sectors to propose easily applicable practical alternatives to facilitate the integration of graduates into the labor market [1]-[3]. Due to the lack of an information technology (IT) system to provide reliable and centralized indicators and statistics on the employability of Moroccan university graduates, higher education in Morocco has been exposed for several years to a set of resulting issues [4]. On the one hand, a lack of reliable data on the issue and the other hand in the face of the global economic crisis, in particular, due to the pandemic [5]-[7]. The high commission for planning (HCP) conducts annual national surveys on employment, but on a diverse and very large population [8].

In addition, the higher council for education, training and scientific research through the national evaluation authority (INE) occasionally carry out surveys on the integration of university graduates in consultation with universities [9]. According to a review of the literature [10]-[13], very few technical research projects have focused on this subject, most of which are carried out by researchers in economic or social studies [14], [15]. Data warehouses are the most significant component of strategic decision-making for business in the last years. This new approach of data analysis which is designed to support managerial decision making has become functional tools used as a repository of information [16]-[18].

The modern recruitment process generally starts online, as indicated by the millions of job ads [19] on recruitment platforms and millions of candidate profiles on the professional social networks [20]. A vast number of approaches have been applied to job offers and candidates profiles [21], with however a low focus
on the skills themselves. Bennett and Landauer [22] proposes to use the topic model latent Dirichlet allocation (LDA) on natural language recruitment data, including job offers. The approach is related to the terminology extraction [23], for which the common approaches are linguistic or statistical, and sometimes involve some machine learning for filtering the irrelevant terms [8]. Despite these different job offers; the unemployment rate of graduates is very high in our society. Talent acquisition and qualified candidates are now the biggest challenges for recruiters and business owners around the world. To help solve this socio-economic problem, this paper proposes a novel approach. We take the case Sidi Mohammed Ben Abdellah University (USMBA) of fez in Morocco, which has realized a new academic data warehouse as a powerful and reliable tool to connect the laureates of the university with the professional world using generated data. Several benefits could be reached by developing a university data warehouse as providing a centralized source of information accessible and enabling administrators to make better decisions based on data available in legacy databases. One of the most valuable points of this system is that it is based on real data and able to play the role of a bridge between the two entities (graduates and recruiters) for given a quality job opportunity.

In the next section, our methodology is described with a brief detail of system architecture used; we explain the design of our scalable smart meter data generator. We also provide the required background on big data concepts and the frameworks for smart grid [24] big data analysis. Section 3 presents the reporting and data publication. Finally, conclusions are given in section 4.

2. RESEARCH METHOD

Due to the lack of a unified and centralized information system at the Moroccan university level, the proposed architecture is based on a data warehouse supplied by several heterogeneous data sources. Consequently, the use of business intelligence and data-mining tools are essential for simple and efficient analysis and exploitation of real-time data. Business Intelligence consists of applications and technologies that help companies to have a wide knowledge about their own business performances. A business intelligence system in the university and recruiting context has wide knowledge about the skills of graduates’ qualified candidates for recruiters. The academic data warehouse is designed to provide a valid tool that satisfies the following needs: a unique system of analysis and reporting, easy access to information on job offers, have statistics to consolidate a new strategy, and improving the quality of work in the professional world [16].

2.1. Academic data framework architecture

As shown in Figure 1 describe the overall architecture of the academic data framework on the typical multi-level. As shown in Figure 1, smart data framework architecture contains three layers, data feed layer, integration and data warehousing layer, and analysis layer. The design of integration and data warehousing is based on web services. In the following paragraph, let us describe our different framework component and their functionalities.
2.2. Framework analysis system model

The academic data framework presents the architecture of our smart grid big data framework and data analysis system model. Our solution can be divided into three blocks: data generator based on data feed, database as an ingestion data warehousing and data analysis. Source databases as data generator contain transactional data. The Figure 1 shows four source databases:

- Apogee: is an integrated management software package developed by the agency for the mutualization of universities and institutions. It is intended for the management of registrations and students’ files.

- LinkedIn and viadeo: two online platforms that mainly used for professional networking and allows job seekers to post their CVs and employers to post jobs. Our framework can connect to both platforms via APIs to retrieve a heterogeneous set of data.

2.2.1. LinkedIn API algorithm

This new framework can connect to LinkedIn data through a “PHP LinkedIn SDK” to fetch company and profile information via API. This API makes possible to fetch profile information like name, email, and updates; fetch company information like name, profile, updates and more. Our algorithm is based on the following script:

- Install the LinkedIn SDK and Client ID: as shown in Figure 2 allows the launch of the LinkedIn SDK client installation with the hypertext preprocessor (PHP) technology and then the definition of a client object with an identification (ID) code and a password key.

Figure 2. LinkedIn SDK and client ID installation algorithm

- Saving the token: the Figure 3 allows the saving of the token through the following steps: add a composer for autoloader, import the library of client class, instantiate the LinkedIn object client, load the token from the file and set the token for client.

Figure 3. Algorithm of saving of the token
The Figure 3 allows the saving of the token through the following steps: add a composer for autoloader, import the library of client class, instantiate the linkedIn object client, load the token from the file and set the token for client:

- Anapec [23]: the national agency for the promotion of employment and skills, it collects job offers from employers and it guides young entrepreneurs for the realization of their economic projects. It is connected through a web service protocol as an xml file to data warehouse with JSON. In the following, the Web service protocol algorithm.

2.2.2. Web service protocol algorithm

The Figure 4 describes the definition of web service protocol communication. With including the config file and then a validation of the algorithm. The algorithm is based on a request method which retrieves the server ID through the HTTP protocol and the server status, then he sends and validates the information through a JSON file to ensure communication between the entities.

```php
<?php
    // Include config.php
    include_once('config.php');
    if($_SERVER['REQUEST_METHOD'] == 'PUT') {
        $uid = isset($_SERVER['HTTP_UID']) ? mysql_real_escape_string($_SERVER['HTTP_UID']) : mysql_real_escape_string($_SERVER['HTTP_UID'])
        // Add your validations
        if(!empty($uid)) {
            $sql = mysql_query("UPDATE `tuts_user` SET `status` = 1, `msg` = 'Status updated!' WHERE `uid` = $uid"); return true;
        } else {
            $json = array("status" => 0, "msg" => "Error updating user ID not defined");
        }
    } else {
        $json = array("status" => 0, "msg" => "User ID not defined");
    }
    mysql_close($conn);
    // Output header */
    header('Content-type: application/json');
    echo json_encode($json);
```

Figure 4. Web service protocol algorithm

2.3. Hadoop Framework

The academic data framework is based on Hadoop open-source framework in our university as a tool to manage the collected data that cannot handled with the traditional management methods. Hadoop is an open-source framework founded by Apache foundation. It is used for running data applications and storing a massive amount of data [25]. Hadoop offers the competence to handle virtually unlimited concurrent jobs or tasks, a massive technique of storage for any type of data and tremendous processing capability. As described in Figure 5, there are many advantages of using Hadoop. Hadoop is one of the tools to manage huge amount of data because it can easily extract information from heterogeneous data. Advantages of using Hadoop as shown in Figure 5.

Communication method between our framework and Hadoop in Figure 6, we present the communication algorithm based on MapReduce. The procedure of our communication protocol algorithm is based on sending map-reduce programs to computers where the actual data resides. During a MapReduce job, Hadoop sends map and reduce tasks to appropriate servers in the cluster. The Figure 6 illustrates the communication protocol based on MapReduce.
Big data based architecture to bringing together graduates and recruiters: … (Abdemounaime Hamdane)

As described in the Table 1, the framework operates on key-value pairs. The key and value classes have to be serializable by the tools and hence; it is required to implement the writable interface. The key classes have a role to implement the interface to facilitate sorting by our framework. MapReduce algorithm in the Figure 7 describes the definition of a function of process units Map by using a mapper class with 2 inputs (key type and value type) and 2 outputs (key type and value type). The Figure 8 describes the definition of a function of reduce and a main function to send and display the results.

\[(\text{Input}) \quad <k1, v1> \rightarrow \text{map} \rightarrow <k2, v2> \rightarrow \text{reduce} \rightarrow <k3, v3> \quad (\text{Output})\]

<table>
<thead>
<tr>
<th>Table 1. MapReduce keys values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Map</td>
</tr>
<tr>
<td>Reduce</td>
</tr>
</tbody>
</table>

![Figure 5. Advantages of using Hadoop](image1)

![Figure 6. communication algorithm based on MapReduce](image2)

![Figure 7. MapReduce algorithm (function of map)](image3)
2.4. Implementation of big data in smart grid FRAMEWORK

As shown in Figure 9, we present the detailed architecture of the proposed solution. Our grid framework retrieves all the data about laureates through the different data sources in order to feed our database as data lake. On the other side, the university has a real database of partner companies with it. All of its data is shared in a company space as Job Offers Data (JOD) that connects the two entities.

To present qualified candidates and guarantees CV to the job market, the academic data framework is based on an extract, transform, load (ETL) process. The ETL process loads data from sources of databases “data lake”, clean it and unify all the data into target tables of data warehouse “final data lake” (verified cv, qualified candidates, effective jobs offers) in order to provide an effective tool as a bridge to connect the two entities laureates and recruiters through university. Figure 9 illustrates the detailed architecture of communication between all components.

Figure 9. Detailed architecture of the proposed solution
This new framework brings great benefit in the recruitment market whether for laureates or for companies through the publication of real data to connect the two entities.

- **Benefits of laureates**: there are, of course, a wide range of benefits for graduates using this framework, including free use of platform services, easy access to information on internship and job offers and improved career opportunities and career counseling.

- **Benefits of recruitment company**: for recruitment company, the benefits includes systematic identification of the characteristics of diplomas. However, the benefits also extend to the possibility of selecting a qualified candidate, whose academic career has been certified by the universities, as well as the free use of platform services and the improvement of the quality of work in the professional world.

- **Benefits of university**: this platform offers great advantages and benefits for the university. It allows easy access to support for the employment of higher education graduates. It provides access to documents and reports for decision-making processes and educational planning. It also makes it possible to produce statistics to consolidate a new strategy aiming a better formation-employment connection.

### 3. RESULTS AND DATA PUBLICATION

Data analyses techniques are used to achieve reports and responses to complex queries. The system can produce statistics as those reported in Tables 2 and 3. The report in Table 2 groups the students curriculum vitae (CV) results between years 2018 and 2019 grouped by academic year and university faculty. The report in Table 3 groups the number of laureates hired in professional world by academic years (2018-2019) and university faculty. For 2020, it is currently being processed.

**Table 2.** Count the students CV results grouped by academic year and faculty

<table>
<thead>
<tr>
<th>Faculty</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sciences Dhar El Mehraz</td>
<td>873</td>
<td>1156</td>
</tr>
<tr>
<td>Sciences and technique</td>
<td>535</td>
<td>645</td>
</tr>
<tr>
<td>Legal science economies and Socials</td>
<td>565</td>
<td>590</td>
</tr>
<tr>
<td>Technology</td>
<td>187</td>
<td>356</td>
</tr>
<tr>
<td>Higher normal school</td>
<td>389</td>
<td>765</td>
</tr>
</tbody>
</table>

Table 2 illustrates the results of the data processed in the process of our framework grid data between the 2018 and 2019. All the results are processed for each faculty in order to have a follow-up of the progress by each one. In 2019, we generated 600 Cvs (verified CV, qualified candidates) on average through our tools which are published in our JOD “Job Offers Data”.

**Table 3.** Count the number of laureates hired in professional world by academic years (2018-2019) and university faculty

<table>
<thead>
<tr>
<th>Faculty</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sciences Dhar El Mehraz</td>
<td>95</td>
<td>123</td>
</tr>
<tr>
<td>Sciences and technique</td>
<td>110</td>
<td>133</td>
</tr>
<tr>
<td>Legal sciences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economies and Socials</td>
<td>96</td>
<td>143</td>
</tr>
<tr>
<td>Technology</td>
<td>67</td>
<td>98</td>
</tr>
</tbody>
</table>

Table 3 illustrates the results of the data of laureates hired in professional world through our Framework grid data between the 2018 and 2019. All the results are processed for each faculty in order to have a follow-up of the progress by each one. In 2019, we generated 100 CVs hired on average through our tools, which are published in our JOD “Job Offers Data”.

### 4. CONCLUSION

In this paper, we proposed a framework to analyze smart grid big data architecture regarding graduates and recruiters. At first, we have presented our architecture of the data grid framework, and then we introduced a scalable data generator to overcome to lack of access to real smart grid big data as job offers data. After we have presented our results and data publication to analyze the evolution of employed graduates into the professional world and make a good bridge between two entities. Future work will provide the extension of this university data framework Layer. This will be done in order to provide approximate more
query processing for complex algorithm applications that allows more speed analytical queries in order to publish real and effective indicators.

ACKNOWLEDGEMENTS
The Sidi Mohamed Ben Abdellah University of Morocco and more particularly by its President and their deans/directors and by establishment’s staff supported this work on this paper.

REFERENCES
BIOGRAPHIES OF AUTHORS

Abdelmounaim Hamdane
his higher studies in National Schools of applied Sciences Fez - Morocco (ENSAF) 2012-2015, Phd student at University Sidi Mohamed Ben Abdellah Fez- Morocco 2015-2022. His served a software engineer in web development and software solutions, also his served as an IT consultant in different management sectors. He can be contacted at email: abdelmounaaim.hamdane@usmba.ac.ma.

Nadir Belhaj Hamdane Phd specialized in big data architectures, AI and blockchain. His higher studies in National Schools of applied Sciences Fez - Morocco (ENSAF) 2012-2015, Phd student at University Sidi Mohamed Ben Abdellah Fez- Morocco 2015-2022. He can be contacted at email: nadir.belhaj@usmba.ac.ma.

Halima El Hamdaoui PhD in Computer science. She is a member of the Laboratory of Artificial Intelligence, Data Sciences, and Emerging Systems. Her research interests include advanced information systems, database systems, big data, machine learning, medical information systems and telemedicine. She can be contacted at email: halima.elhamdaoui@gmail.com.

Karima AISSAOUI Ph.D., Ing. Professor, Faculty of Sciences, University Med 1st, Oujda – Morocco. Her research focused on: Artificial intelligence, E-learning, Testing, Computer Security, Software Development. She can be contacted at email: k.aissaoui@ump.ma.

Moulhime El Bekkali is a professor in the National School of Applied Sciences at the Sidi Mohamed Ben Abdellah University (USMBA), Fez, Morocco. He is a member of the Laboratory of Artificial Intelligence, Data Sciences, and Emerging systems. He can be contacted at email: moulhime.elbekkali@usmba.ac.ma.

Prof. Dr. Nour El houda Chaoui is a full Professor in Computer Science at Sidi Mohamed Ben Abedellah University (USMBA), in the National School of Engineers (ENSAF) – Fez – Morocco. She is a former Professor of Computer Science at Ibn Zohr University in National School of Engineers (ENSA)- Agadir-Morocco. Dr. Chaoui teaches courses in Business Intelligence, Database administration, education and leadership. She is a member of the Laboratory of Artificial Intelligence, Data Sciences, and Emerging Systems. Her research interests include advanced information systems, database systems, big data, machine learning, medical information systems and telemedicine. She can be contacted at email: houda.chaoui@usmba.ac.ma.

Big data based architecture to bringing together graduates and recruiters: … (Abdemounaime Hamdane)