Teaching the Basic Commands of NoSQL Databases Using Neo4j in Vocational Education and Training (VET)

Dimitrios Kotsifakos, Dimitrios Magetos, Alexandros Veletsos and Christos Douligeris

Abstract—This paper presents a new pedagogical approach to the teaching of the Databases course in a Vocational Education and Training (VET) environment. We propose an exemplary teaching scenario for basic database commands using Neo4j. This approach is necessitated by the fact that many online/network environments are taught in VET without any special preparation in the corresponding laboratories. Furthermore, our article describes various technical aspects of the NoSQL software family in general and of Neo4j in particular.

Index Terms—VET Education, Laboratory Practices, NoSQL.

I. INTRODUCTION

Many changes and reforms have been launched recently in Vocational Education Training (VET) in Greece with the goal to substantially upgrade the quality of VET [1]. A Strategic Framework [2], which is based on a holistic view of the entire VET system, was completed in 2016. The implementation of these reforms focuses on the recognition and strengthening of the social role of VET. These reforms also aim to keep VET in line with the requirements of the Greek society and economy, providing quality education to the pupils and students who choose it and, finally, ensuring decent professional outlets for VET graduates.

For students or learners who have chosen courses or specializations related to Informatics in the formal vocational secondary education (EPAL) and in the Vocational Training Institutes (IEK), the subject of DataBases (DB) is required by the respective curricula. The curricula of the Second Section of Informatics and Computer Science in the 3rd grade of Informatics of VET (Informatics Technician and Software Application Technician) explicitly mention the teaching of DB. By avoiding the vertical specialization in the VET courses, the students are expected to make an effort to acquire basic and often time-consuming knowledge about Database Management Systems (DBMS) environments.

As mentioned in the curriculum, "the overall aim of the course is for the learner to be provided with all the knowledge that should allow him/her to recognize and distinguish the advantages of the applications developed in a DBMS environment and to be able to implement simple applications in such an environment". Since, nowadays, the students are fully familiar with online environments as users, the question of VET and education as a whole is how we can think creatively as producers. To do this we need to expand the knowledge in the general and growing field of Data Science. Specifically, for the course "Databases Management Systems and their Applications on the Internet" of the 3rd Class of the VET’s High School, a certain number of hours (theory and laboratory) are foreseen according to the course curriculum. The students who follow the Specialization «Informatics Technician and Software Application Technician» of VET should acquire not only a basic knowledge about operating systems but they should also learn how to acquire a basic knowledge, skills and expertise in Databases. In addition, they should learn how to implement simple Internet applications in combination with Databases for organizing, managing and processing information (Sheet No. 2010, No Φ2 / 141426 / Δ4, September 16, 2015, page 23815, Maroussi).

The relational databases (RDBs) have been established and used in the labor market to store data for decades with SQL used as the language of interaction [3]. Nevertheless, today, there are many large scale applications on the Internet which save and process a huge amount of data. NoSQL systems have been proposed as more efficient and effective in saving and indexing these big groups of data. At the same time, they are able to efficiently respond to multiple types of user requests.

The technology of NoSQL databases was developed by large Internet companies, such as Google, Facebook, Amazon and LinkedIn. A large number of companies have adopted NoSQL databases for a variety of use cases. This proliferation has been led by Big Users and Big Data and the increase in the use of the Internet and of Cloud Computing [21, 22]. In parallel, it has been observed that the number of amazingly talented and well-trained graduates has increased exponentially. Even though these graduates were keen on holding job positions which were related to Data Management, they were lacking in vital parts of knowledge about how the data are managed because the knowledge gained in the traditional school education covers only a small part of what exists in real work conditions [23]. As a consequence, when they enter the job market they face difficulties in using their skills and knowledge they have gained from their curricula.

This article attempts to contribute in closing this gap and, particularly, to support new Computer Scientists to be able to use their knowledge and skills effectively and successfully. At the same time, as far as the knowledge transmission is concerned, it tries to support the VET
teachers who are involved in the related subjects. This article presents alternative ways of teaching the database lessons that can be used in VET. The new curricula of the Informatics specialties exploit and study environments such as Neo4j (https://neo4j.com/) which are encountered when using NoSQL, without any special preparation or training in the classroom. By applying a specific and well-focused teaching scenario, the gap we mentioned before can be covered to a large extent. Moreover, the contribution of our research concerns the prospects of linking this knowledge to its immediate professional use, as it works not only in the apprenticeship classes but also for the newcomers in the IT job market.

Our contribution to the relevant research and the pedagogical feasibility of our approach is revealed through the development of a proposed teaching scenario. The scenario focuses on the teaching of basic NoSQL databases using Neo4j. Neo4j is one of the best-known graph databases that helps to find relationships between data and to organize a complete picture of them. The purpose of the teaching scenario is to make students understand and use the web development environment of a widespread NoSQL database system, Neo4j, combined with the use of basic node and relationship creation commands. Our teaching scenario highlights the main advantages of using NoSQL databases and suggests how an introductory course should be organized. The curriculum we propose can be part of the curriculum of the Technical Applications of Computer Science and the Software Applications of the 3rd class of VET’s High School. It is also a part of the 3rd class of evening VET classes in the field of Informatics and particularly in the context of teaching the course "Databases Management Systems (DBMS) and their Applications on the Internet". The teaching scenario presented in the following section addresses "Databases on the Internet".

This paper is organized as follows: in Section 2, we present an introduction to the subject under study and we describe the identified gap in the research. We make clear the scientific basis of our proposed work and show its relevance: the demand of DBMS and the basic principles of NoSQL. In Section 3.1, we discuss the technical requirements for our scenario and in Section 3.2 the teaching methodology. We also describe in detail the steps students have to follow in order to be able to interpret the basic commands that they should be taught. Section 4 presents the implementation and the evaluation of the proposed scenario, the functionality of the teaching, and some qualitative features of the implementation. The final section concludes with future work in the area.

II. INTRODUCTION TO NO SQL DATABASES; THE IDENTIFIED GAP IN THE RESEARCH

In the last decade, the majority of the applications are developed on the Internet [14]. So far, most of the requirements of these applications have been adequately covered by relational databases (RDBs) [5]. Although databases were designed to meet market needs - and they have performed adequately to date - they can no longer handle the sizes managed by modern applications, and they are also unable to store and display large data sets [9, 10] that can be found in applications such as Big Data Analytics, Business Intelligence and Social Networking which produce data sizes of peta-bytes. These applications, which are often characterized as Large Scale Applications (LSA), need to store and process large amounts of data. Moreover, they are used by large user populations, which may be globally scattered, and are particularly demanding in terms of the response time of these systems. Thus, they face various difficulties in coping with the required agility challenges [6, 7] and they cannot easily benefit from the cheap storage and processing power which are available today [8].

The effort which was made in order for these needs to be covered has created and brought to the fore a new set of databases with different characteristics than the already known RDBs. This database group, was NoSQL, includes a variety of different technologies developed in response to these new requirements. The main feature of NoSQLdatabases is that they do not follow the RDBMS (Relational Database Management System) model which, as mentioned before, has been traditionally used in most cases previously [17]. A strict relational schema is usually an obstacle to network/web applications, which must handle texts, comments, photos, videos, source code and other information saved in multi-tables in order to be accessed directly. Furthermore, NoSQL systems are designed for saving and processing big data in a large number of servers [11]. Some of the well-known NoSQL databases are Google Bigtable [12], and HBase open source applications [13].

III. SUGGESTED SCENARIO METHODOLOGY

A. Technical requirements

Our scenario is in line with the updated database teaching suggestions [20] and could be a good starting point for those who are interested in teaching modern database examples.

Our proposed teaching scenario requires a fixed internet connection. This is the only technical requirement to teach the basic principles of NoSQL databases. NoSQL databases do not use any kind of structured system such as: tables which are structured units in a common relational system. Moreover, they do not use a Structured Query Language (SQL) for data manipulation [15]. The data are organized and saved in a completely non-relational so improved way that they are attached and retrieved easily [16]. For our scenario we should use Neo4j, which uses Cypher [18] as the language for queries. Cypher is a declarative query language for batch databases.

Fig. 1. The declarative query language Cypher
The Cypher property graph query language is an evolving language, originally designed and implemented as part of the Neo4j graph database, and it is currently used by several commercial database products and researchers [19]. Fig. 1 shows the relationship between a Cypher query and a JSON response. Cypher queries are very expressive, in the sense that they closely resemble the way humans intuitively understand graphs. Because it is declarative, it focuses on clearly defining what to draw from the graph and not how to get it. Finally, Neo4j is a property graph database. This means that beyond attributes and labels, each relationship has an initial and a final node as well as a direction. Many times, however, in an application, we do not want relationships to be directional. In this case, instead of bidirectional relationships, we can simply ignore the direction of relationships during queries.

For the teaching of the NEO4J databases using graphs we can use the web application http://console.neo4j.org/, which is a free development environment for NoSQL graph databases. In addition, using this application we are able to create simple as well as complex data charts and nodes connected with links. Furthermore, we can access information which is stored on the nodes of charts.

B. Teaching approach

The proposed scenario can be taught in the 3rd class of VET in the framework of teaching the section "Databases on the Internet". The needs of the course can be thoroughly covered by the equipment of the Laboratory Center (LC) that supports the EPAL. In the beginning of the lesson the teacher he must describe how the basic commands for NoSQL databases must be organized using Neo4j. In the next step the teacher should explain the basic graph environmental use functions of Cypher. Also, should explain the commands that are appropriate for creating a small chart in a Neo4j environment and also must explain the basic DBMS functions are applied. The expected duration of the scenario is two (2) hours. The scenario targets the creation of simple and multiple nodes together with the relationships among them.

The development and completion of the scenario requires some prior knowledge of database management. The students should have understood the basic concepts of databases. They should have also understood and implemented ways of designing a database. Finally, they have to know how to design, develop and maintain a database application. Having repeated the previous lessons briefly, the teacher should try to provoke the students' interest. In addition, the students should be familiar with the Neo4j online environment (Fig 2) and the basic environmental use functions. The overall presentation of the image of the environment is necessary for the student's orientation. In VET teachers organize knowledge by giving a general picture of the exercise and then taking part in individual applications.

After completing the scenario, the students should be able to retrieve, store and create simple commands using Neo4j. The students should work individually or in groups of two or three (2-3) students (depending on the equipment availability at each Laboratory Centre). The teacher notifies the beginning and the direction of the activities, either by leaflets or by using an overhead projector. The students, through their worksheets and activities, should try to deepen their understanding and get familiarized with the NoSQL concept using the “http://console.neo4j.org/” environment. The teacher should participate actively in the learning process by: monitoring and controlling the student responses during the exercise, and by encouraging them to continue their exploratory effort through targeted questions and concerns. The teacher's help can be supportive, cooperative or even instructive to students if needed. The teacher's help is organized in such way because students learn interacting with web environment. Students must learn how managed real problems about db and acquire correct decision-making criteria for these issues. Help can be given explicitly and sometimes off-the-record. The students should not devote time to design applications but that their effort should be focused on the implementation of the described activity. Since the Neo4j browser is a command driven client, like a web-based shell environment, it is perfect for running ad-hoc graph queries, with just enough ability to prototype a Neo4j-based application. The students should remain focused in writing and running graph queries with Cypher. Exportable tabular results of any query result, graph visualization of query results containing nodes. Finally visualized relationships and convenient exploration represented of Neo4j's REST API.

C. Commands in details

The version of a Neo4j database can be displayed using the store-info command of neo4j-admin (Fig 3).

```
neo4j-admin store-info --store=(path-to-store)
```

Fig. 3. The store-info command

Another basic function is the backups store information (Fig 4). The Neo4j Admin “backups-info” command can be used while another backup of Neo4j is present. The output tells us that the backup is made with the standard store format, which all editions of Neo4j support. Our scenario uses the newest version of the format known to Neo4j. This means that no format migration is necessary for the backup to be restored. The teacher must explain the basic commands for the creation of a chart that may represent the relationships among a writer’s two or three books.

```
$neo4j/home/bin/neo4j-admin info --store=backup/graph.db.2018-11-11
```

Fig. 4. The store-info store back ups command

In the next stage, the teacher should ask the students to
use the appropriate commands to create a graph. During the lesson, the teacher can also ask the students to create a node in Neo4j by simply identifying the name of the node to be created through the CREATE command (Fig 5).

```
CREATE (node name);
```

Fig. 5. The CREATE command

In order to verify the creation of the node, a return query must be executed at the application command line. This query returns all nodes in the database. This results in the created node being displayed. The CREATE command of Neo4j is also used to create multiple nodes simultaneously (Fig 6).

```
CREATE (node1), (node2);
```

Fig. 6. The CREATE command for multiple nodes

The teacher must explain to the students that in Neo4j, a relationship is an element with which we associate two nodes of a graph. To create a relationship, we can use the CREATE command (Fig 7).

```
CREATE (node1)-[:RELATIONSHIP]->(node2);
```

Fig. 7. The CREATE command for a relationship

The module of the course should be completed by displaying the commands needed to create and display two nodes. Initially, the teacher should ask the students to create two nodes, e.g. "Gre" and "Jim" in the database (Fig 8). Then, the teacher should ask the relationship between these two nodes to be called BATSMAN_OF (Fig 9). After that, the teacher should represent the created relationship (Fig 10).

```
CREATE (Jim:player(name: "Jim Magetos"), YOB: 1907, POB: "Sparti")
CREATE (Gre:CountrY (name: "Greece")
```

Fig. 8. The CREATE command for two nodes

```
CREATE (Jim:player)-[:BATSMAN_OF]->(Gre)
```

Fig. 9. The CREATE command for a relationship

Upon completion of the above exercises, the teacher must evaluate how many of the previous students have understood the basic commands ask questions. Then the teacher should close the first teaching hour. In the second hour, the teacher should explain the new activities and should propose appropriate consolidation exercises. At the beginning of the second hour, the teacher should point out to students that all the popular web applications use databases to store user information. For instance, Facebook has a database where its information about each member, their friends’ list, publications, etc is stored. The Amazon online store has a database where it stores the products that someone can buy. Google maintains in a database the information contained in each web page that is posted on the Internet. This can be done by linking the subject matter to the students’ daily routine.

The teacher should ask the students to think about how the applications we use on the Internet, such as an online store, store various pieces of information whose operation in a database is necessary. Moreover, the students can be helped by being told that when the number of users, the number of transactions, and, more generally, the volume of operations produced by thousands of users at the same time over the Internet is too large then the traditional databases do not offer satisfactory processing times.

```
Fig. 10. Representation of the created relationship
```

IV. DISCUSSION - QUALITATIVE ASSESSMENT

Our article described various new technical aspects of the NoSQL software family and of Neo4j in particular. All the new tools for DBMS are cognitive tools that do not simply reinforce their users’ cognitive system but reorganize it by suggesting and associating new ways of thinking. The improvement of teaching databases requires an understanding of the relevant representations and cognitive student models about DBMS as well as the difficulties they encounter in designing simple as well as more complex systems. So far, the research which concerns the teaching of databases combined with the use of web environments focuses mainly on the type of the taught structures, the chosen environment, the programming language used and the teaching approach followed. In our scenario, we propose a new tool to update database teaching. The advantages of our scenario are:

- The command management interface does not need to be installed. It is activated directly with individual accounts on the Internet.
- The use of graphic commands provides a more appealing approach to students.
- The students feel that they come in contact with modern environments and are getting prepared to manage large databases as the ones encountered in social media applications.

The difficulty in consolidating the management skills of DBMS on the part of the students lies on the fact that they treat them as users rather than as designers. This results in misunderstandings about the handling of DBMS. In any case, the teacher should take into account the students’ previous knowledge (ideas, perceptions, representations), bearing in mind that the pre-scientific pupils’ previous knowledge and attitudes are not easily eliminated, but they

DOI: http://dx.doi.org/10.24018/ejers.2019.0.CIE.1291
constitute important cognitive barriers to building new knowledge. The students by working together collaborate to learn how to manage big problems together. This gives them the courage to face ever more difficult problems. However, it has been documented widely in the literature that by using exploratory teaching methods and by managing the net/web environments, the students develop their own self-action and their difficulties are overcome.

The next point of the discussion concerns the implementation of the course on web platforms directly and not on preinstalled software on a local server. For the modern teaching, as shown in [20], the installation of the environments, which are now used by the labor market on a local server, is not mandatory. Indeed, for the Neo4j as well as for other NoSQL software the installation on a local server is not necessary. The only thing we need for teaching the lesson is a secure and stable connection to the Internet. The additional educational value of our proposal lies on the development of the students’ skills in VET and the corresponding specialties of IEK on the issues of managing the current internet reality. The students must act as producers and not as consumers. The guidelines we propose are a confirmation of the fact that the curricula in cutting-edge technologies cannot be statically addressed but should be organized in a spirit of constant renewal and in accordance with a specific implementation horizon combined with the relevant social standards of pedagogical effectiveness.

For this analysis, it was our goal to upgrade the employment prospects of VET and IEK graduates and to ensure their smooth and safe entry into the labor market with the aim of reconciling vocational education with employment and the prospect of supporting them after they receive their degree. The Database Management System (DBMS) overview and our suggestions in conjunction with the proposed scenario, included in the above framework, offer an up-to-date immediate useful solution for the teaching and implementation of databases. The web environments we suggest enable not only to teach the basic concepts of NoSQL databases but also to make VET graduates able to handle a variety of objects in real working conditions since they should be able to handle flexible web applications that the underlying databases support.

V. CONCLUSIONS AND FUTURE PROPOSALS

This article described several aspects to be integrated in an immediate and professional use for teaching new technical aspects of the NoSQL software family, Neo4j in particular, through the presentation of an introductory scenario. As databases are ubiquitous, our proposal can be used not only by VET; DBMS affect many disciplines beyond computer science and business. Moreover, our proposal can be further extended in a VET environment. The students of many specialties in VET can benefit from an introduction to fundamental database concepts that goes beyond information fluency [24]. For VET students, an introductory database course typically relies on computer science for prerequisite courses that cover programming, data structures, algorithms, and discrete mathematical structures [4]. According to all of the above, the VET’s graduates will be able to manage jobs related to Data Management and possess knowledge about how to manage the data. These associations, combined with knowledge acquired in traditional school education, will cover much of what exists in real working conditions. Our next work concerns this direction and the linking of practice with more and more modern data processing tools.

ACKNOWLEDGMENT

This work has been supported by the University of Piraeus Research Center and the Greek Secretariat for Research and Technology/National Contributions/Project Sauron.

REFERENCES


DOI: http://dx.doi.org/10.24018/ejers.2019.0.CIE.1291
Dimitrios Kotsifakos is a PhD student in the Department of Informatics at the University of Piraeus, awarded with Distinction (2010), and an undergraduate degree in Electronic Engineering from the Technological Educational Institute (A-TEI) of Athens (1990). In 2012, he successfully completed a one-year Pedagogical Training Program at the School of Pedagogical and Technological Education (ASPETE) in Athens. He has participated in seminars, training programs related to Electronics, Informatics and Electrical Studies as well as in international and Greek conferences. He is appointed as a Secondary School Teacher of Electronics (Vocational Education and Training (VET)). He has also completed his studies in Advanced Music Theory and Composition and holds a Certificate in Fugue and Guitar. He has participated in international seminars and conferences on musical composition and performance. He is especially active in innovative actions concerning the educational process and he has received awards for the cultural programs he has developed. He is the father of four children.

Dimitrios S. Magetos received his MSc (2016) in Advanced Computing and Informatics Systems from the Department of Informatics, University of Piraeus, Greece. He was born in 1967 in Sparta and received his BS (1994) from the Department of Informatics of the Athens University of Economics and Business. He has been working for twenty-four years in the public education as an informatics teacher. His research interests include data management, cloud computing and web-based information systems. He is a member of the Greek Computer Society.

Alexandros Veletsos is an MSc candidate in the Department of Informatics at the University of Piraeus. He holds a BS in Physics from the National and Kapodistrian University of Athens. He has taught high school students physics for two years (2015 – 2017) voluntarily and he has worked as a management employee (11/11/2016 – 21/12/2016).

Christos Douligeris, currently a professor at the Department of Informatics, University of Piraeus, Greece, held positions with the Department of Electrical and Computer Engineering at the University of Miami. He was an associate member of the Hellenic Authority for Information and Communication Assurance and Privacy and the President and CEO Hellenic Electronic Governance for Social Security SA. Dr. Douligeris has published extensively in the networking scientific literature and he has participated in many research and development projects. His main research interests lie in the areas of computer networking, communications, network security, cyber security, web science, data analytics, new technologies in education and emergency response operations.