TEACHING DATABASE ADMINISTRATION IN THE WORLD OF BIG DATA AND SMALL BUDGETS

Tatiana Simmonds
University of Greenwich
Park Row
London SE10 9LS
T.Simmonds@gre.ac.uk
www.cms.gre.ac.uk

ABSTRACT

Teaching advanced database subjects such as database administration within a university environment traditionally has many challenges; however, the modern world of big data and engineered database systems, as well as increased pressure on universities to enhance graduates' employability creates new sets of demands on delivering and assessing learning in both practical and theoretical aspects of database administration courses. This paper discusses the issues involved in teaching Oracle database administration courses for postgraduate students in the School of Computing and Mathematical Sciences at the University of Greenwich. The paper offers various solutions and innovative teaching methods that have proved to be successful in addressing modern industry requirements.

Keywords

Database administration, Oracle, big data, Exadata, employability, database curriculum.

1. INTRODUCTION

Database administration is usually understood as a wide range of tasks for managing and maintaining database management systems (DBMS), starting with the database design and finishing with backup and recovery as shown in Figure 1.

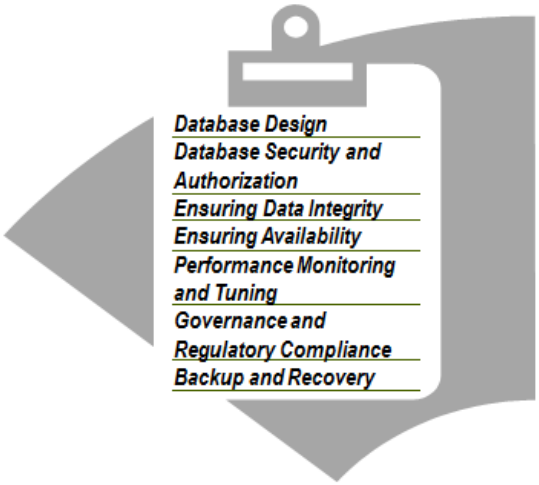

Figure 1: Database administration tasks

Some textbooks include a much more substantial list of topics, providing a comprehensive listing of tasks and responsibilities for the database administrator (DBA); see Figure 2.

Figure 2: Database administration areas covered in textbooks (for example, [2])

In universities, database administration is usually considered to be one of the advanced database subjects and is often included in the curriculum at postgraduate level; topics such as database design, query processing and some aspects of database security may be part of undergraduate courses. This paper will concentrate on topics specifically related to database management and administration.

2. CHALLENGES

Teaching database administration within a university environment has traditionally had many challenges and the modern world has added a few new problems.

2.1 The University's constraints

The majority of our postgraduate courses at the University of Greenwich are designed as 15-credit modules that run for one semester. They consist of 12 lectures, tutorials and laboratory exercises. In addition, for our postgraduate programmes we have two intakes: September starters and January starters, any subject must therefore fit into the 12 lectures and cannot be spread over two or more modules. This presents a challenge for database administration, since, as shown above, this subject is quite expansive. So we had to make a careful selection of the areas to be taught based on learning outcomes of the course, as well as considering employers' input and job market demands.

Also, while general information and theoretical concepts of database administration can be covered in the lectures, practical aspects such as managing database instances, creating and removing users, creating and managing tablespaces and especially database backup and recovery are difficult to demonstrate to students and even more difficult for students to practise in an environment where one database is shared by many students and each student account has limited permissions.

2.2 Employability and job markets

Employability has been a hot topic in HE discussions for quite some time and recently it has become even hotter. The Higher Education Funding Council for England (HEFCE) 2010 mandate required "all higher education institutions to publish employability statements on the help they provide to students to improve their employability and transition into work." [1]

There are many definitions of employability and many aspects to it. For the purpose of this paper we will use the definition provided by the Bologna Process (European level process implemented in 47 countries, which define the European Higher Education Area (EHEA)). It identifies employability as "the ability to gain initial employment, to maintain employment, and to be able to move around within the labour market." [2] The role of

higher education in this context is therefore “to equip students with skills and attributes (knowledge, attitudes and behaviours) that individuals need in the workplace and that employers require, and to ensure that people have the opportunities to maintain or renew those skills and attributes throughout their working lives. At the end of a course, students will thus have an in-depth knowledge of their subject as well as generic employability skills.” [2]

Our University targets employability as a key objective and the Strategic Plan for the next five years proposed by our Vice Chancellor aims to increase student employability KPI, as measured by the Destinations of Leavers from Higher Education (DLHE) survey from 83% to 90%, and – more importantly – it specifically outlines “the shift from what students know to what they can do on graduation.” [3]

Now the question is: do we have a clear picture of what the graduate should be able to do in order to get a job within 6 months after graduation? The relationship between universities and employers has been a subject of research and debate for quite some time now. [4] But one of the outcomes produced by the employability working group of the aforementioned Bologna process in 2008 showed that “employers do not think that universities are doing enough to prepare graduates for the world of work.” [5]

The current economic situation makes employers very demanding, especially in the world of database administration jobs. If 10 years ago many employers would hire university graduates with the minimum Oracle database administration knowledge (as long as they had a solid background in relational databases) and then train them in all necessary aspects of Oracle database administration (companies had huge training and professional education budgets), these days things are different. Those budgets for training are long gone: companies cannot afford to ‘mould’ the graduate into a perfect database administrator that suits their needs; they want to hire a perfect database administrator even at a junior level.

This therefore puts even more pressure on universities to improve employability for their graduates and equip them with the skills that increase their chances to get a job in a very sparse and competitive job market.

2.2.1 Is it worth it to be a DBA?

We have to understand the employment market while defining specific skill sets associated with potential graduates. It would be useful to find answers to the following questions:

- What do employers require for database administration-related jobs?

  DBA jobs are usually vendor-specific and would require different skills sets for different databases; for Oracle DBA, current employers usually require the following skills/experience: capability in all Oracle administration; experience with Oracle 11g on both Microsoft and UNIX platforms; practical experience of Oracle database backup and recovery, including RMAN; competence in shell scripting and PL/SQL programming; experience of database tuning and session diagnostics, etc.

  In the past few years employers have often added migrating large volume Oracle databases to new platforms and exposure to Oracle Exadata as necessary skills. We also found that for SQL Server or MySQL administrators, knowledge of Oracle database administration is usually considered to be a big advantage.

- What are young people (and maybe even not so young people), potential employees, looking for? Many online database forums and discussions give us an insight into this world. Here are a few useful comments from one of the forums [12]:

  “I am in Oracle database administration for 3 years and honestly speaking I have just scratched the tip of iceberg, there are lots and lots of stuff to learn, but believe me it’s challenging, exciting and eventually rewarding both for soul, body and pocket.”

  “Schools don't have a formal DBA education path. Most DBAs got there by transitioning from some other role in IT”.

- Is it worth to be a DBA?

  An IT Salary Survey 2012 conducted by the IT media company TechTarget showed that the average annual salary for a database administrator in the US is $91,523 (compared to an IT manager’s annual salary of $93,640, or a CIO’s annual salary of $126,827) [6]. A 2013 survey conducted by UK-based company Randstad Technologies revealed that a database administrator’s annual salary can vary between £39,600 and £48,400 for a junior to £59,400–£72,600 for a senior DBA, while for a data warehouse manager it can start from £57,600 at junior level and go up to £105,600 at senior level. [7]

2.3 New technologies

2.3.1 Big data

The term 'big data' has many definitions; for the purpose of this paper we will use one of the common understandings of this phenomenon as *data plus analytics,* where 'data' is usually a large amount of complex, unstructured data and 'analytics' are algorithms applied to that data in order to extract useful information. The preferable way to store these large amounts of unstructured data is to use so-called NoSQL storage types or NoSQL databases. After a few years of excitement surrounding NoSQL databases, now the common understanding is that they do not replace relational databases, but rather complement them. So the teaching of database-related subjects should include exposure to unstructured and schemaless data stores, as discussed in [13]. One of the challenges, though, lies in the variety of NoSQL datastore types – key/value-based, document-based, column/family-based, graph-based, multi-model and others. Another challenge is in the number of their implementations (there are currently 150) including open-source and commercial releases.

Big data has had an evolutionary effect in the area of database administration, requiring in addition to traditional DBA expertise a new range of skill sets, including: integrating various types of data stores into a data centre using distributed and clustered environments; ability to move data between relational databases and NoSQL databases; as well as knowledge of data analysis systems such as Hadoop. Moreover, the conventional separation between the roles of system administrator, database administrator and developer in the world of big data becomes blurred, as working with systems such as Hadoop will require knowledge of distributed file systems as well as using programming languages such as Java or Python to write analytical algorithms.

2.3.2 Engineered Systems

In 2008 Oracle introduced the Exadata database machine – the first of the Engineered Systems, three years after IBM introduced PureSystems. This was the beginning of a new wave in database technology development and datacentre computing. Engineered systems (sometimes also called 'integrated systems') combine software and hardware together in one 'out-of-box' system wiring together processors, storage and network connections. These systems are considered to be a core for constantly growing data centres and a good solution for many big data problems.

However, at the same time, these new innovative systems have introduced new demands on database administrators to "undergo a shift in thinking as well as attain new skills" [3]. The database administration domain more often becomes data centre administration and, besides traditional database administration areas of expertise, requires knowledge of network, storage and other hardware-related topics as shown in Figure 3.

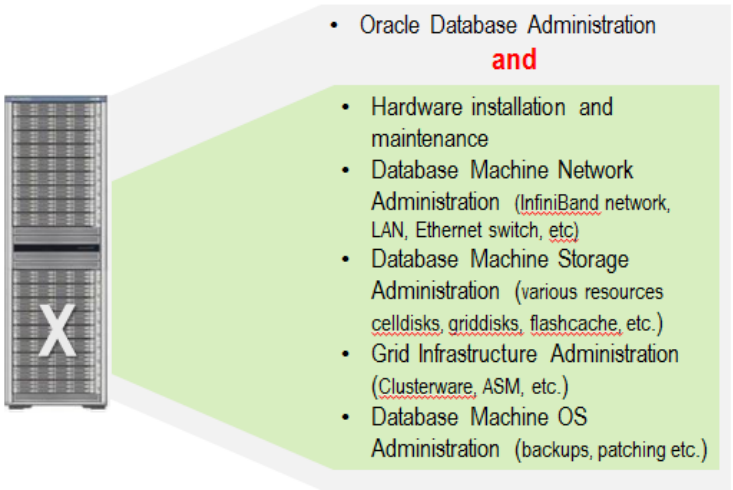

Figure 3: Oracle Exadata database machine skills

3. TEACHING DATABASE ADMINISTRATION

3.1 Using Virtual Machines

The database administration course has been running in our school at postgraduate level for a few years now. As it is impossible to cover all aspects of database administration in 12 lectures and labs/tutorials, we have

chosen the most crucial areas as shown in Figure 4 and require students to complement the contact hours with the tutor with 2–3 hours a week of independent self-study using textbooks suggested in the reading list, as well as extensive online resources and database forums.

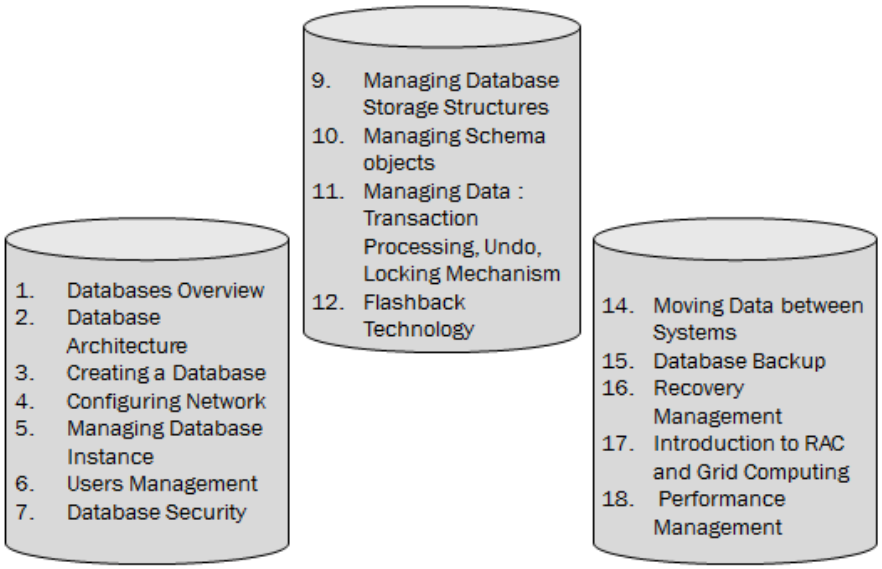

Figure 4: Database administration topics

In order that students acquire the necessary practical skills in database administration, we provide each student with a Virtual Machine (VM) as his/her personal lab environment. Every year we have had, on average, 30 VMs created for each run of the course. VMs were created using VM Ware and run on a central server – our 'Classroom environment' ESX cluster – which includes two physical servers with 16 cores each (4 processors with 4 cores per socket) and 128 GB RAM. The students access their VMs via a browser from any lab on campus or from home via VPN. There have been no major issues with workload or performance running this setup.

Each VM was configured as 2 GB RAM, 1CPU (Xeon® 2.13 GHz) and 20 GB hard drive; Microsoft Windows Server 2003 SP2 was used as the operating system. We did consider running the course on Linux OS, but since the majority of the students were not familiar with this operating system we decided to use Windows Server. In future we intend to include a core course dedicated to the Linux operating system in the database-related programmes and therefore would be able to switch the database administration course to Linux OS. The latest release of Oracle Enterprise Edition database 11g Release 2 software was installed on each VM.

The labs were designed as a progressive sequence of exercises that required the students to complete the previous lab/tutorial before continuing with the next, starting from creating a database based on specified criteria. The main learning objectives for laboratory exercises were focused on problem-solving aspects of database administration. The students had to assume different roles – database administrators as well as end users – in order to simulate situations as close as possible to real-life scenarios. As part of the assessment for learning of the practical skills, the students had to upload a weekly report containing explanations of problems or errors encountered during the lab, their understanding of what might cause those errors and the actions they had to take in order to solve the problem.

For example, as part of the lab dedicated to managing tablespaces, the students were required to create a tablespace according to provided specifications (which intentionally defined a small size for that tablespace and no auto extent) and make it available to the end users. Then, switching to the end-user role, the students had to create a table in the above-mentioned tablespace and start populating it with a large number or rows. At some point they received a message in the SQL*Plus window stating "unable to extend the table ...". They then had to find out the reason for the error and take action to solve it.

In some labs erroneous situations were created for students and they had to make more investigations in order to find the reason for the problem. For example, the lab for Flashback Technologies simulated a real-life scenario where an end user calls a database administrator saying he/she cannot access their regular table. The students had to identify the root of the problem and solve it. After checking user privileges and other possible causes, they found that the user table had been accidentally dropped; they therefore had to recover it using an appropriate Flashback functionality.

3.2 Facing new technologies

It is difficult to fit database administration of relational databases and NoSQL databases into one course/module, so we cover NoSQL databases in a separate course on big data. However, it is useful to introduce the idea of unstructured data storage and make necessary connections to big data when teaching certain aspects such as External Tables or Real Application Clusters.

Some topics that are important for any engineered solution administrator were already part of the course – the use of External Tables, RMAN Backups, and Introduction to RAC and Clusterware; other aspects related to network and storage are covered in separate modules, and we amended the structure of our postgraduate programme for database specialists to include mandatory courses on network architecture and management, storage and Linux OS.

In order to provide the students with exposure to new technologies of Engineered Systems, we invited a guest lecturer from Oracle Corp. to give a talk about the Oracle Exadata database machine; the talk was open not only to students on a database administration course, but also to other postgraduate and undergraduate students. It was very informative and the students had the chance to ask questions related not only to Exadata, but also to other aspects of the database administrator's job. In future we are planning to organise an excursion to Oracle headquarters in Reading to see the actual Exadata machine at work.

3.3 Assessments and results

The students on the course were assessed using two units of assessment: written examination and coursework consisting of two parts: a logbook of weekly uploads and a final report. Each upload was marked on a weekly basis in order to provide students with timely and specific formative feedback. The examination included questions to test deep understanding of the concepts, as well as the ability to solve problems and to work under pressure.

The results were very satisfying, with the pass rate increasing every year from 82.4% in 2011 to 88% in 2012 and the student survey consistently showing 100% of students satisfied with the quality of the course.

3.3.1 Students' feedback

The student survey also revealed that students enjoyed the course and found it useful and helpful. Here are some comments:

- *"To be honest, of all the courses this is one of the toughest subject, but it was made simple, understandable and more interesting to the students";*
- *"This was the best lecture and learning process I had all sessions";*
- *"Database Students Should not miss such lectures";*
- *"Lecture hand-outs, lab exercises were well designed; teaching standards and explanation for student's questions was excellent. It should be core course for Database related programmes".*

But the most important feedback we received from the graduates, reflecting on how this course helped with their employability, was:

"Working with some of the latest technologies available, this course gave me the knowledge and practical experience I needed to improve and drive the management of our Database systems forward, making use of the vast functionality available to us, brought to surface by this course. We support four enterprise-level Oracle databases and a number of other DBMS, which I am happy to say, are now in the best position they have been in and incorporate a vast amount of information learnt from this course."

"This programme nurtures students in such a way that they will be highly competitive in the job market, evidenced by my ability to secure a job offer as a Database Developer in a leading US based Market Research Company even before completing the degree."

4. CONCLUSIONS

There are several large influences that shape the way we teach database administration in the university environment today. On the one hand companies' budgets for training are in decline; therefore, they want to recruit employees who already have the specific DBA skill set. On the other hand universities are under pressure to improve students' employability and equip them with the skills that increase their chances in a sparse and competitive job market.

In addition to this, the subject of databases is going through a revolutionary stage and places more DBA skills on the required list.

As discussed in this paper, there are various creative ways to adapt. The use of Virtual Machines can solve some problems with the centralised university environment and provide a perfect platform for students to experience full control of the database and acquire necessary practical skills. Close collaboration with employers and industry experts provide the students with insights; professional certification short courses enable the students to acquire better employment and advancement opportunities. The timely introduction of new courses can cover current trends and cutting-edge technologies.

5. REFERENCES

[1] Mullins C. (2012) Database Administration: The Complete Guide to Practices and Procedures. Second Edition.

[2] Universities to publish employability statements. HEFCE News 2010.
http://www.hefce.ac.uk/news/newsarchive/2010/name,62266,en.html

[3] Employability. The official Bologna Process website July 2007 - June 2010
http://www.ond.vlaanderen.be/hogeronderwijs/bologna/actionlines/employability.htm

[4] University of Greenwich Strategic Plan 2012–17.Making Greenwich Great.
http://www.gre.ac.uk/__data/assets/pdf_file/0006/747951/University-of-Greenwich-Strategic-Plan-2012-2017V2.pdf

[5] Yorke M. (2006) Employability in higher education: what it is – what it is not.
http://www.heacademy.ac.uk/assets/documents/employability/id116_employability_in_higher_education_336.pdf

[6] The employability working group update prepared for the Bologna Seminar in Luxembourg 2008. http://www.ond.vlaanderen.be/hogeronderwijs/bologna/BolognaSeminars/documents/Luxembourg/Luxembourg_Seminar_Employability_WG_update.pdf

[7] TechTarget's annual IT Salary survey 2012 results.
http://cdn.ttgtmedia.com/searchDataCenter/downloads/IT_Salary_Survey_2012_final.pdf

[8] The 2013 IT Salary Survey. Randstad Technologies.
http://www.randstad.co.uk/technologies/career-tools/salary-survey-2013/ accessed 14 June 2013

[9] Kanaracus, C., "Oracle's Exadata Means Changes for DBAs" PCWorld Oct 4, 2011

[10] Oracle Exadata X3 Administration Certification Exam.
http://education.oracle.com/pls/web_prod-plq-dad/db_pages.getpage?page_id=5001&get_params=p_exam_id:1Z0-027 accessed 14 June 2013

[11] Computer Weekly European User Awards: Storage.
http://www.computerweekly.com/EuropeanUserAwardsStorage accessed 16 June 2013

[12] Is it worth to be DBA?/Oracle Forums
https://forums.oracle.com/forums/thread.jspa?threadID=325448 accessed 18 June 2013

[13] Stainer C. Introducing NoSQL into Database Curriculum. Proceedings TLAD 2012.