EXPERIENCES IN THE DEVELOPMENT OF BIG DATA FOR ENTERPRISE MANAGEMENT COURSE

Ricardo Chalmeta, Manuel Mollar, Sergio Orenga

Grupo de Integración y Re-ingeniería de Sistemas. Dept. Lenguajes y Sistemas Informáticos. Universitat Jaume I (SPAIN)

Abstract

The Universitat Jaume I (UJI) of Castellón, Spain, is teaching the Master's Degree in Intelligent Systems. The main objective is to cover the great lack of consistent and updated Intelligent Systems skills in the university graduate students. It involves perception, learning and automated decision processes, robotic manipulation, data storage and processing, contextual information, visualisation and interactive environments.

The master is composed of 20 courses. One of them addresses the development and implementation of a Big Data ecosystem in an organization. Big Data has become a key factor for organizations to gain a competitive advantage over competitors because it can help to bring a lot of benefits to organizations, such as to develop new products or services; make better, quick and automated strategic and operational decisions; or identify what was happened, and predict what will happen in the immediate future.

The big data for enterprise management course equips students with the necessary basic knowledge and competencies they need to design and utilize Big Data for enterprise management, considering not only technological aspects, but also the management of organization policies and people.

The aim of this paper is to describe the course objectives, the target competencies, the course contents, the assessments, and how the teaching methodology and the resources are used to teach the subject.

Keywords: Intelligent Systems, big data, enterprise management, learning experiences.

1   INTRODUCION

Advances in infrastructure, storage techniques, and data-processing allow huge volumes of structured and unstructured customer data to be analysed in a very fast and efficient way, and with an acceptable cost for most organisations. Due to the amount and complexity of these data, it is difficult to process them using traditional tools. So the use of Big Data technology is essential in order to take advantage of this kind of data [1].

To describe Big Data, the Vs frameworks have emerged as a common structure. Hence, Big Data can be defined as follows: "Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" [2]. Three other Vs have been added, subsequently, to the previous ones: Veracity, which represents the unreliability inherent in some sources of data; Variability (and complexity), which refers to the fact that Big Data are generated through a myriad of sources; and Value, which refers to the fact that data in their original form usually have a low value relative to their volume, but a high value can be obtained by analysing large volumes of such data [3].

Big Data has the potential to help companies improve operations and make faster, more intelligent decisions. This data, when captured, formatted, manipulated, stored, and analyzed can help a company to gain useful insight to increase revenues, get or retain customers, and improve operations.

This paper describes the big data for enterprise management course of the Master's Degree in Intelligent Systems at the Universitat Jaume I (UJI) of Castellón, Spain. It shows the target competencies, the course contents, the assessments, and how the teaching methodology and the resources are used to teach the subject.

2   OBJECTIVES

The objectives of this subject are to (1) introduce students the concept and challenge of big data (6 V's), and (2) teach students in applying skills and tools to manage and analyze the big data.The course gives students an in-depth understanding of a wide range of fundamental algorithms and processing platforms used in big data management.

The course covers Big Data Fundamentals, including the characteristics of Big Data, the sources Big Data (such as social media, sensor data, and geospatial data), as well as the challenges imposed around information management, data analytics, privacy and security, as well as platforms and architectures. Emphasis is given to non-relational databases by examining techniques for storing and processing large volumes of structured and unstructured data, streaming data as well as complex analytics on them. Data warehouses will also be presented as a solution to handling big data and business intelligence applications.

3   LEARNING OUTCOMES/CAPABILITY DEVELOPMENT

3.1   Specific capabilities

- To know What Big Data is
- Big Data technology platforms beyond the data warehouse
- How to analyse un-modelled, multi-structured data using Hadoop, MapReduce & Spark
- To know How to integrate Big Data with traditional data warehouses and BI systems
- To use Big Data systems and technology to support decision making
- To identify business and technical requirements for a Big Data solution
- To apply relevant theories, concepts and techniques to solve real-world Big Data problems
- To perform big data analyses
- To visualize the results of big data analyses
- To design and build Big Data applications based on users' needs
- To increase technical skills through the designing and building of business tools or applications to resolve specific business problems.

3.2   Cognitive Capabilities

- Computer literacy
- Analytic/design (conduct research, analyse, reason, solve problems, interpret data)
- Appreciative (evaluate ideas, make judgements, think critically, multidisciplinary perspectives, self-recognition)

3.3   Behavioural Capabilities:

- Personal (think and act independently, creativity, tolerant of ambiguity, flexible)
- Interpersonal (listening, presentation, communicate, negotiate, group dynamics, collaboration)

4   LEARNING METHODOLOGY

Learning methodology combines classical lectures and Active Learning. Lectures are used to transmit information from the teacher to students simultaneously about the course contents. Lectures are a good way of providing an outline or overview of a subject of study.

To avoid the criticism of the lecture which is seen as a potentially quite passive and therefore not a particularly effective learning experience for students, active learning is used too. Active Learning involves students directly and engages them actively in the learning process itself. Students are involved in all stages of planning, design, execution and evaluation.

Active Teaching Methods used in the course are: Group projects, Questionnaires, Practical work, Brainstorming, Group discussion, Case studies, Sharing your own experiences, Group work, Role play, Individual students – class presentation

Active Teaching Methods used Out of School are: Surveys, Market research, Questionnaire design

5 SYLLABUS/COURSE CONTENT

5.1 Lectures

1 Introduction to Big Data: The 6 V's, their challenges and application domains.

2 Collection of Big Data: Eventual Consistency and NoSQL systems MongoDB, Google BigTable

3 Large-Scale Data Analytics Systems: Auto-Parallel Data Programming; MapReduce, Hive, and Parallel Databases

4 Basic Statistical Analysis: Fruad and Benfords Law, Bayesian Introduction, Heteroskedasticity

5 Machine Learning Systems for Big Data

6 Graph Analytics: Graph structures (diameter, connectivity, centrality), PageRank, Triangle counting

7 Sentiment Analysis Last update in July 2014

8 Data Visualization: Data types and dimensions; Visual encoding and perception

5.2 Active Learning

The main objective of the active learning is that students apply the W2KM methodology [4] to an enterprise. The W2KM provide practical guidance to manage the entire project life cycle by indicating the steps that are needed to execute the development and implementation of Big Data ecosystems in any organization. They have to do it in groups composed by 5 students. The W2KM methodology is organised in phases and outlines the activities to be performed, the techniques and supporting tools to be used, and the expected results for each phase [5].

So, the students have to develop the following activities of a Big Data development:

5.2.1 Phase 1. Draft

The aim of this phase is to study the feasibility of the project for the organisation, that is to say, whether it would be in the organisation's interests to undertake it and if it is going to be cost effective.

5.2.2 Phase 2. Planning

Several tasks are done in this phase: to achieve the commitment of Management; to set up the project management team; to construct and moderate between the communities that exist within the organisation; to establish specific aims by using SWOT analysis and mechanisms of control; and to determine the material and human resources that will be needed to carry out the project.

5.2.3 Phase 3. Analysis

The first thing to be undertaken in this phase is to identify the target knowledge, that is, all the knowledge that the company wants to know about because it is useful to the organisation and will provide an added value when utilised. The next stage is to identify the input variables that will make it possible to obtain the target knowledge inside each knowledge block, and their sources.

5.2.4 Phase 4. Design

This is the phase in which the functional, technological and graphic design of the KMS 2.0 is carried out. First of all, the functional design of the Web 2.0 and Big Data tools is defined. For each Web 2.0 and Big Data tool that is going to be used in the KMS 2.0 it is necessary to specify the way in which the input variables are going to be managed in order to obtain the target knowledge of each conceptual block of knowledge. After, the technological design is then carried out. To do so, first the characteristics of the Web 2.0 and Big Data tools that are going to be implemented must be defined. After that, a decision must be made as to whether the software will be custom built or if the

(commercial or free-distribution) application will be acquired and later tailored to meet specific needs. The next step is to define the hardware requirements.

5.2.5 Phase 5. Development

In this phase, the Web 2.0 and Big Data tools are installed, developed/customised and tested, and the corresponding user manuals are produced.

5.2.6 Phase 6. Implementation

In the implementation phase, the system is put into operation. At first the system is only used by a subset of the final users (called "key users") of each Web 2.0 and Big Data tool. The purpose is to take advantage of their own experiences or impressions to debug and refine them. The key users of a tool are the people in charge of implementing, customising and debugging it.

5.2.7 Phase 7. Control

The control phase spans the whole useful life of the system. In this phase the system is monitored and, if necessary, capabilities are adapted or modified to solve errors and improve the system or adapt it to new ways of working in the organisation.

6   ASSESSMENT

Active learning work: 50%

Lectures: 50%

7   CONCLUSIONS

This paper shows the curriculum design of a big data for enterprise management course. This curriculum has been developed by three teachers of computer systems department at Universitat Jaume I. The course is organized in eight blocks.

Students are very satisfied with course contents and organization. The course combines lectures with practical work, so theory can be understood easier, and students obtain new knowledge and new skills, similar to professional work.

REFERENCES

[1]   Syed A, Gillela K, Venugopal C (2013) The Future Revolution on Big Data. International Journal of Advanced Research in Computer and Communication Engineering 2(6):2446-2451.

[2]   Gartner IT Glossary (2016). Big Data. Available at: http://www.gartner.com/it-glossary/big-data/ (accessed 25 June 2016)

[3]   Gandomi A, Haider M (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35(2):137-144

[4]   Orenga-Roglá, S. and Chalmeta, R. 2016. "Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies." SpringerPlus (2016) 5:1462. DOI 10.1186/s40064-016-3128-y.

[5]   Orenga-Roglá, S. and Chalmeta, R. 2017. "Methodology for the implementation of knowledge management systems 2.0: A case study in an oil and gas company." In press, accepted manuscript in Business & Information Systems Engineering (BISE).