BIG DATA MANAGEMENT AND ANALYTICS

1. GENERAL

SCHOOL			
ACADEMIC UNIT			
LEVEL OF STUDIES	ΠΡΟΠΤΥΧΙΑΚΟ		
COURSE CODE	ICE-8103	**SEMESTER**	8th
COURSE TITLE	Big Data Management and Analytics		
INDEPENDENT TEACHING ACTIVITIES if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits		**WEEKLY TEACHING HOURS**	**CREDITS**
Lectures		3	
Tutorials		1	
		4	**5**
COURSE TYPE general background, special background, specialised general knowledge, skills development	Επιστημονικής Περιοχής, Ανάπτυξης Δεξιοτήτων **Εμβάθυνσης/Ειδίκευσης**		
PREREQUISITE COURSES:	Suggested prerequisites: “Design and implementation of algorithms” and “Data Mining”.		
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	English		
IS THE COURSE OFFERED TO ERASMUS STUDENTS	No		
COURSE WEBSITE (URL)			

2. LEARNING OUTCOMES

Learning outcomes

The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described.

Consult Appendix A

- Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area
- Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B
- Guidelines for writing Learning Outcomes

The aim of the course is to present large-scale data management techniques and advanced Data Mining issues, as well as their applications.

Upon completion of the courses the students will be able to:

• Understand issues related to large-scale data repositories (HDMS etc.),

• Delve into methods and practices related to large-scale Data Mining on the World Wide Web and cloud-based systems,

• Familiarize themselves with research approaches and new solutions to the problems that arise,

• Familiarize themselves with applications of theory to real problems, in order to acquire specialized problem-solving skills, which are required in research and/or innovation in order to develop new knowledge and processes as well as to integrate knowledge from different fields,

• Gain critical awareness of knowledge issues in the field of large-scale data management and its interconnection with other fields,

- Acquire the necessary skills that will allow them to continue their studies in the field of Large-Scale Data Management and Analytics, in a autonomous fashion, to a large extent.

General Competencies

Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?

Search for, analysis and synthesis of data and information, with the use of the necessary technology	Project planning and management Respect for difference and multiculturalism Respect for the natural environment
Adapting to new situations	Showing social, professional and ethical responsibility and sensitivity to gender issues
Decision-making	Criticism and self-criticism
Working independently	Production of free, creative and inductive thinking
Team work	……
Working in an international environment	Others…
Working in an interdisciplinary environment	…….
Production of new research ideas	

- Search, analysis and synthesis of data and information, using the necessary technologies,
- Individual work,
- Work in an interdisciplinary environment,
- Production of new research ideas,
- Creative and critical thinking,

3. SYLLABUS

- Large-scale file systems, the Map-Reduce and Spark platforms.
- Link Analysis,
- Advertising on the World Wide Web,
- Data Mining from Social Network Graphs,
- Recommender Systems,
- Link-open-data (LOD) platforms,
- Big data and the Semantic Web,
- Data Mining and Business Intelligence,

4. TEACHING and LEARNING METHODS - EVALUATION

DELIVERY Face-to-face, Distance learning, etc.	- (a) Face to face. In-class lectures with active student participation. There will be presentations and exerices on the course web-page as well as in class. - (b) Tutorials and project work.
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY Use of ICT in teaching, laboratory education, communication with students	• Use of specialized software for big data management an analysis. • Use of the university's online teaching platform for posting theory and exercises. • Use of e-mail and the online teaching

<table>
<tr><td></td><td>platgorm for communication with the students.</td></tr>
<tr><td>TEACHING METHODS
The manner and methods of teaching are described in detail. Lectures, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.

The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS</td><td>
<table>
<tr><th>Acitivity</th><th>Semester workload</th></tr>
<tr><td>Lectures</td><td>39</td></tr>
<tr><td>Tutorials</td><td>13</td></tr>
<tr><td>Project/Exercises</td><td>25</td></tr>
<tr><td>Individual study</td><td>48</td></tr>
<tr><td>Total Course Hours
(25 ώρες φόρτου εργασίας ανά πιστωτική μονάδα)</td><td>125</td></tr>
</table>
</td></tr>
<tr><td>STUDENT PERFORMANCE EVALUATION
Description of the evaluation procedure

Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other

Specifically-defined evaluation criteria are given, and if and where they are accessible to students.</td><td>I. Final written-exam (80%)

II. Individual exercises (20%)</td></tr>
</table>

5. ATTACHED BILBIOGRAPHY

- Suggested bilbiography :

1. J. Leskovec, A. Rajaraman, J.D. Ullman, Mining of Massive Datasets, Cambridge, 2nd edition, 2016. <- I have used this extensively and highly recommend it !
2. P. Tan, M. Steinbach, V. Kumar, Introdcution to Data Mining, Pearson, 2 edition, 2018.
3. S. Walkowiac, Big Data Analytics with R, Packt Publishing, 2016.
4. P. Cimiano, O. Corcho, V. Presutti, L. Hollink, S. Rudolph (eds.), "The Semantic Web: Semantics and Big Data," Proceedings of *10th International Conference, ESWC 2013*, Montpellier, LNCS 7882, Springer, 2013.
5. H. Chen, R. H. L. Chiang, V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly*, vol. 36, issue 4, pp.1165-1188, December 2012.
6. W. Fan, A. Bifet, "Mining Big Data: Current Status, and Forecast to the

Future," *SIGKDD Explorations*, vol.14, issue 2, 2014.
7. A. R. Ganguly and A. Gupta, *Data Mining Technologies and Decision Support Systems for Business and Scientific Applications*, Encyclopedia of Data Warehousing and Mining, 2005.
8. R. Kohavi, N. J. Rothleder, E. Simoudis, "Emerging trends in business analytics," *Communications of the ACM - Evolving data mining into solutions for insights*, vol. 45, issue 8, pp 45-48, August 2002.
9. J.P. Shim, M. Warkentin, J.F. Courtney, D.J. Power, R. Sharda, Ch. Carlsson, "Past, Present and Future of Decision Support Technology", *Decision Support Systems: Directions for the Next Decade*, vol.33, issue 2, pp. 111-126, June 2002.
10. Y. Sun, J. Han, *Mining Heterogeneous Information Networks: Principles and Methodologies*, Morgan & Claypool, 2012.
11. H.J. Watson, B. H. Wixom, "The Current State of Business Intelligence," *IEEE Computer*, vol. 40, issue 9, pp. 96-99, September 2007.
12. I. Witten, E. Frank, M. Hall, *Data Mining: Practical Machine Learning Tools and Techniques* (4th edition), Morgan Kaufmann, 2019.
13. M. d'Aquin, G. Kronberger, M. Suárez-Figueroa, "Combining Data Mining and Ontology Engineering to enrich Ontologies and Linked Data", Proceedings of *1st International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data*, Heraklion, 2012.
14. W. Fan, A. Bifet, "Mining Big Data: Current Status, and Forecast to the Future," *SIGKDD Explorations*, vol.14, issue 2, 2014.
15. T. Heath, C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool, 2011.
16. L. Palathingal, S. Dascalu, F. C. Harris Jr, Y. Varol, "A Brief Survey of Data Curation Literature", Proceedings of *CATA 2015*, Honolulu, Hawaii, March 2015,
17. H. Paulheim, "Exploiting Linked Open Data as Background Knowledge in Data Mining", Proceedings of *International Workshop on Data Mining on Linked Data (DMoLD)*, Prague, 2013.
18. J.P. Shim, M. Warkentin, J.F. Courtney, D.J. Power, R. Sharda, Ch. Carlsson, "Past, Present and Future of Decision Support Technology," Decision Support Systems: Directions for the Next Decade, vol.33, issue 2, pp. 111-126, June 2002.
19. X. Wu et al., "Top 10 Algorithms in Data Mining," *Knowledge and Information Systems*, vol. 14, pp. 1-37, 2008.
20. E. Turban, R. Sharda, D. Delen, D. King, *Business Intelligence: A Managerial Approach* (2nd Edition), Prentice Hall, 2011.
21. M. North, *Data Mining for the Masses with implementations in RapidMiner and R*, 2016.
22. H.J. Watson, B. H. Wixom, "The Current State of Business Intelligence," *IEEE Computer*, vol. 40, issue 9, , pp 96-99, September 2007.