COURSE OUTLINE

1. Information about the program

1.1 Higher education institution	**“Alexandru Ioan Cuza” University of Iaşi**
1.2 Faculty	**Faculty of Economics and Business Administration**
1.3 Departament	**Department of Accounting, Information Systems and Statistics**
1.4 Field of study	**Business Informatics**
1.5 Cycle of study	**Master**
1.6 Study program / Qualification	**Software Development and Business Information Systems**

2. Information about the course

2.1 Course title			**Big Data Persistence and Processing** (Course proposed via the project eUAIC - cod F-PNRR-GDU-1-2022-0109 î to replace the Database Logic in Business Applications course)				
2.2 Course coordinator			**Lect. Ionut Hrubaru, Ph.D.**				
2.3 Seminar coordinator			**Ionut Hrubaru, Lucian Lazar**				
2.4 Year of study	**I**	**2.5** Semester	**I**	**2.6** Type of evaluation*	**P**	**2.7** Course status**	**C**

* *MT-MID-TERM, O-ORAL EXAM, E-EXAM, M-MIXED; ** C-compulsory/O-optional/E-elective*

3. Estimated time allocation (hours per semester and teaching activities)

3.1 Number of hours per week	**4**	out of which: **3.2** course	**2**	**3.3** seminar / laboratory	**2**
3.4 Total number of hours per semester	**56**	out of which: **3.5** course	**28**	**3.6** seminar / laboratory	**28**

Time allocation	h
Study based on course book, course materials, bibliography and other	**30**
Supplementary study in the library, on electronic platforms and on the field	**14**
Preparing seminars/laboratories, assignments, papers, portfolios and essays	**30**
Tutorship	**14**
Examination	**6**
Other activities	

3.7 Total hours of individual study	**94**
3.8 Total hours per semester	**150**
3.9 Number of credits	**6**

4. Prerequisites (if applicable)

4.1 Referring to curriculum	**Databases (or similar)**
4.2 Referring to competences	**Not applicable**

5. Conditions (if applicable)

5.1 For the course	- Video projector, Microsoft Teams, and Moodle (for ad-hoc and scheduled tests) - Distributed PostgreSQL Citus cluster on eUAIC and RaaS-IS platforms for demonstrations and case studies. - Distributed MongoDB cluster on eUAIC and RaaS-IS platforms for demonstrations and case studies"
5.2 For the seminar / laboratory	- Oracle DB Server 12c (or newer) or PostgreSQL 14 (or newer) installed on one's own computer. - Oracle DB Server 12c (or newer) or PostgreSQL 14 (or newer) installed on the RaaS-IS and eUAIC platforms (for demonstrations and project-based assessments) - Distributed PostgreSQL Citus cluster on eUAIC and RaaS-IS platforms (for demonstrations and project-based assessments) - MongoDB Server 6 (or newer) installed on the computer • Distributed MongoDB cluster on eUAIC and RaaS-IS platforms (for demonstrations and project-based assessments)

6. Specific competences accumulated

Professional competencies	- C1.5 Development of analysis, design, implementation and testing of an information system based on real-world case studies and compare various solutions (0.5 credits) - C2.1 Mastering theoretical and technological knowledge and tools concerning business data modeling, query, processing, administration and analysis, including Big Data (0.5 credits) - C2.2 Selection and refinement of the methods and techniques for data modeling, persistence, query and analysis, according to the nature of problems and available resources (0.5 credits) - C2.3 Assess the degree of information integrity and validity for organizational data; find the appropriate tools for administration and analysis of business data (0.5 credits) - C2.4 Design the most appropriate solutions for gathering, storage, processing, administration and analysis of business data according to the organizational resources and constraints (0.5 credits) - C2.5 Develop projects and case-studies concerning modeling, implementation (database logic), administration and analysis of data for real-world applications (1.5 credits) - C4.5 Write the specifications and deploy the modules regarding data, applications and services integration (1.5 credits)
Transversal competencies	- CT1 – The ability to communicate and collaborate in teams of different professionals (0.5 credits)

7. Course objectives (based on specific competencies accumulated)

7.1 General objective	- To provide the core knowledge, methodologies and tools in order to deal with modeling and managing of high volumes of business data
7.2 Specific objectives	- Knowledge of the main data storage (persistence) technologies - Ability to design and process stored data (using SQL and procedural extensions of SQL) in large-scale relational databases, exemplified by Oracle or PostgreSQL - Ability to create and operationalize distributed databases using Citus (PostgreSQL) - Developing essential skills in modeling, storing, and distributed processing of information stored in document databases, exemplified by MongoDB. - Basic understanding of the Hadoop ecosystem for data storage and processing using Apache Hive. - Ability to design database schema.

8. Content

8.1	**Course**	**Teaching methods**	**Observations** (time and bibliography)
1.	Big Data. Platforms and formats for data persistence. Data processing languages. Trends.	PPT presentation, discussion,	1 hour [Foster et al., 2021] (cap.4 și 5) [Fotache, 2023]
2.	What is the database logic in an economic application? Minimalism and maximalism in the	PPT presentation, discussion.	1 hour [Fotache, 2023]

	database logic/layer		
3.	Advanced data processing options in SQL. - Oracle dialects.	PPT presentation, script execution, discussion	2 hours [Feuerstein & Pribyl 2014] [Fotache 2009] [Fotache 2023]
4.	Procedural extensions of SQL. Implications for software development	PPT presentation, code execution, discussion.	1 hour [Feuerstein & Pribyl 2014] [Fotache 2023] [Fotache 2009]
5.	Designing the database schema in real applications. Temporal validity of the schema. Normalization vs. performance. Denormalization."	PPT presentation, E-R diagrams, discussion. Case studies	3 hours [Fotache, 2005] [Ponniah, 2007] [Blaha, 2010] [Fotache, 2023]
6.	Stored procedures in Oracle PL/SQL	PPT presentation, diagrams discussion. Case studies	2 hours [Fotache 2023] [Fotache 2009]
7.	Automating processing and organizational rules (Business Rules) using triggers in Oracle PL/SQL	PPT presentation, code execution, discussion. Case studies	2 hours [Fotache 2023] [Fotache 2009]
8.	Open-source Big Data (distributed) architectures for SQL servers. Case study: Citus (PostgreSQL)	PPT presentation, code execution, discussion. Case studies	2 hours Fotache 2023]
9.	Analysis of SQL query performance in Citus, based on the parameters of the distributed setup	PPT presentation, code execution, discussion. Case study	2 hours [Fotache 2023]
10.	Document databases. MongoDB. Modeling, simple queries. Migration from relational databases to document databases	PPT presentation, code execution, discussion. Case study	3 hours [MongoDB, 2022a] [MongoDB, 2022c] [MongoDB, 2022d] [Copeland, 2013] [Fotache et. al, 2023]
11.	Querying MongoDB databases using the Aggregation Framework	PPT presentation, code execution, discussion.	3 hours [MongoDB, 2022b] [MongoDB, 2022c] [Fotache et. al, 2023]
12.	Open-source Big Data (distributed) architectures for document databases. Case study: MongoDB	PPT presentation, code execution, discussion.	2 hours [Fotache 2023]
13.	Analiza performanței interogărilor MongoDB, în funcție de parametrii setupului distribuit.	PPT presentation, code execution, discussion.	2 hours [Fotache et. al, 2023]
14.	Introductory elements regarding the storage and processing of Big Data in the Hadoop ecosystem and data processing using Hive	PPT presentation, code execution, discussion.	2 hours [Fotache, 2023] [White, 2015] [Capriolo et. al., 2012] [Hrubaru ,2016]

Bibliography

References:

Blaha, M. – Patterns of Data Modeling, CRC Press, Boca Raton CA, 2010
Capriolo E., Wampler D., Rutherglen J. (2012). Programming Hive, O'Reilly
Copeland R. (2013). MongoDB Applied Design Patterns, O'Reilly
Feuerstein, S., Pribyl, B. - Oracle PL/SQL Programming, Sixth Edition, O'Reilly, 2014
Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F., Lane, J. (2021). Big Data and Social Science. Data Science Methods and Tools for Research and Practice, 2nd ed., Routhledge, 2021, disponibilă gratuit la: https://textbook.coleridgeinitiative.org
Fotache, M., Strîmbei, C., Creţu, L. (2003). Oracle 9i2. Ghidul dezvoltării aplicaţiilor profesionale, Ed. Polirom, Iaşi, 2003. Disponibilă gratuit la: https://github.com/marinfotache/Database-Logic-in-Business-Applications/tree/master/Oracle%209i2.%20Ghidul%20dezvoltarii%20aplicatiilor%202003
Fotache, M. (2005). Proiectarea bazelor de date. Normalizare şi postnormalizare. Implementări SQL şi Oracle, Ed. Polirom, Iaşi, 2005. Disponibilă gratuit la: https://github.com/marinfotache/Database-Logic-in-Business-Applications/tree/master/Proiectarea%20bazelor%20de%20date.%20Normalizare%20si%20postnormalizare%202005
Fotache, M. (2009). SQL. Dialecte DB2, Oracle, PostgreSQL şi SQL Server, Ed. Polirom, Iaşi, 2009. Disponibilă gratuit la: https://github.com/marinfotache/Baze-de-date-I/tree/master/SQL.%20Dialecte%20DB2-%20Oracle-%20PostgreSQL%20si%20SQL%20Server
Fotache, M. (2023). Logica bazelor de date în aplicații pentru afaceri, UAIC, FEAA, Iaşi, (pagina cursului pe portalul FEAA, Moodle și GitHub - https://github.com/marinfotache/Database-Logic-in-Business-Applications)
MongoDB (2022a). MongoDB Basics. Curs gratuit disponibil furnizat de MongoDB University: https://university.mongodb.com/courses/M001/about
MongoDB (2022b). The MongoDB Aggregation Framework. Curs gratuit disponibil furnizat de MongoDB University: https://university.mongodb.com/courses/M121/about
MongoDB (2022c). MongoDB for SQL Pros. Curs gratuit disponibil furnizat de MongoDB University: https://university.mongodb.com/courses/M100/about
MongoDB (2022d). Data Modeling. Curs gratuit disponibil furnizat de MongoDB University: https://university.mongodb.com/courses/M320/about
Ponniah, P. (2007). Data Modeling Fundamentals, Wiley
White T. (2015). Hadoop. The Definitive Guide (4th Edition), O'Reilly

Compulsory reading:

Throughout the course, other references will also be provided via the FEAA Portal/GitHub platform

Optional reading:

8.2	**Seminar / Laboratory**	**Teaching methods**	**Observations** (time and bibliography)
1.	Installation and configuration issues for Oracle Database Server and Oracle SQL Developer.	Scripts and code execution	2 hours
2.	Creating, updating, and querying Oracle databases using the SQL language	Discussion. SQL solutions writing and analysis	4 hours
3.	IT1 – Individual Test 1 SQL	Moodle platform	2 hours
4	TA1 - team assessment no 1. SQL	Presentation of the team solution	1 hour
5.	Problems and solutions in PL/SQL (1)	Scripts and code execution	4 hours
6.	P11 - Presentation of the 1st part of the project - a real world case database schema	Presentation of the team solution – E-R diagram. Feedback suggestions for	2 hours

		improvement	
7.	Triggers in PL/SQL. Examples. Case studies	Demonstration, scripts, and code execution	2 hours
8.	P12 - Presentation of second part of the project - database logic implemented in Oracle PL/SQL	Applications requirements and diagram discussion, feedback, assessment	2 hours
9.	Setting up a distributed setup for MongoDB on the RaaS-IS and eUAIC platforms	Demonstration, scripts, and code execution	2 hours
10.	Creating and modifying MongoDB databases. Queries using the Aggregation Framework	Discussion, Scripts and code execution	4 hours
11.	IT2 – Individual Test 2 - MongoDB Aggregation Framework	Moodle platform	1 hour
12.	P2 – MongoDB project presentation	Schema presentation, demonstration, scripts and code execution. Discussion	2 hours
13.	Creating a distributed Hadoop solution on the RaaS-IS and eUAIC platforms	Demonstration, scripts, and code execution	2 hours
14.			

Bibliography

The one indicated for the course, plus the references mentioned in the presentations and scripts on the portal.

9. Bridging course content with the expectations of the community, professional associations and representative employers in the field of the program

- The content of this discipline has been decided upon by taking into account both the curricula of some prestigious Western Universities and the demands of the economic environment provided by potential employers, either in the public or in the private IT companies

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Allocation to the final grade (%)
IT1 – Individual Test 1 SQL	Solution validity	Moodle test	15%
TA1 - team assessment no 1. SQL	Data validity and relevance. SQL queries syntax validity	SQL Scripts	15%
P11- 1st part of the project - database schema in Oracle Data Modeler for a real world case)	Validity, relevance and complexity of proposed database schema	Presentation and discussion of the application requirements and E-R diagram with each team member	15%
P12 - The second part of	Validity, relevance,	Presentation and	15%

the project - database logic implemented in Oracle PL/SQL	complexity, and finesse of the solution	discussion of the PL/SQL modules with each team member	
IT2 – Individual Test 2 - MongoDB Aggregation Framework	Solution validity	Moodle test	15%
P2 – MongoDB project presentation	Validity, relevance, complexity, and finesse of the solution	Database schema verification MongoDB Scripts /Aggregation Framework	15%
Correctness of solutions to ad-hoc problems presented during the course andlaboratory.	Validity, relevance, complexity, and finesse of the solutions		10%
10.6 Minimal performance standard			
Design and implement diagrams, scripts and code modules for database logic			

Date	Course coordinator	Seminar coordinator
11.09.2023	Lect. Ionut Hrubaru. Ph.D.	Lect. Ionut Hrubaru, Ph.D. Lucian Lazar

Date of approval in the departament
27.09.2023

Head of Departament
Prof.dr. Mircea Asandului