SYLLABUS

1. Data about the program of study

1.1 Institution	The Technical University of Cluj-Napoca
1.2 Faculty	Faculty of Automation and Computer Science
1.3 Department	Computer Science
1.4 Field of study	Computer Science and Information Technology
1.5 Cycle of study	Master
1.6 Program of study / Qualification	Data Science / Master
1.7 Form of education	Full time

2. Data about the subject

2.1 Subject name	***Data Analytics Pipeline***			Subject code	**8.10**
2.2 Course responsible / lecturer	Prof. dr. eng. Sebestyen Gheorghe - Gheorghe.Sebestyen@cs.utcluj.ro				
2.3 Teachers in charge of seminars / Laboratory / project	Prof. dr. eng. Anca Hângan - anca.hangan@cs.utcluj.ro				
2.4 Year of study	I	2.5 Semester	2	2.6 Type of assessment (E - exam, C - colloquium, V – verification)	E
2.7 Subject category	Formative category: DA – advanced, DS – speciality, DC – complementary				DA
	Optionality: DI – imposed, DO – optional (alternative), DF – optional (free choice)				DO

3. Estimated total time

3.1 Number of hours per week	3	of which:	Course	1	Seminars	1	Laboratory	1	Project	-
3.2 Number of hours per semester	52	of which:	Course	14	Seminars	14	Laboratory	14	Project	-
3.3 Individual study:										
(a) Manual, lecture material and notes, bibliography										15
(b) Supplementary study in the library, online and in the field										15
(c) Preparation for seminars/laboratory works, homework, reports, portfolios, essays										15
(d) Tutoring										10
(e) Exams and tests										3
(f) Other activities:										

3.4 Total hours of individual study (suma (3.3(a)…3.3(f)))	58
3.5 Total hours per semester (3.2+3.4)	100
3.6 Number of credit points	5

4. Pre-requisites (where appropriate)

4.1 Curriculum	Introduction to Big Data, Machine Learning, Distributed System - bachelor
4.2 Competence	Operating with fundamental computer science concepts

5. Requirements (where appropriate)

5.1. For the course	Blackboard, Projector, PC MS Teams Platform
5.2. For the applications	PC, Specific Software

6. Specific competence

6.1 Professional competences	analyse big data analyse business processes build predictive models create data models debug software define technical requirements develop software prototype develop with cloud services perform data cleansing perform data mining perform scientific research provide technical documentation use data processing techniques use software design patterns use software libraries utilise computer-aided software engineering tools utilise machine learning
6.2 Cross competences	develop an analytical approach taking a proactive approach developing strategies to solve problems being open minded coordinate engineering teams

7. Expected Learning Outcomes

Knowledge	The student has knowledge of: • cloud technologies • computer science • data analytics • data models • data storage • data warehouse • database management systems (DBMS) • digital data processing • unstructured data • statistics • ICT debugging tools • computer programming • software components • software libraries • service-oriented modelling

Skills	The student is able to: • create data sets • develop data processing applications • establish data processes • manage ICT data architecture • manage data • manage quantitative data • manage research data • perform dimensionality reduction • process data • use data processing techniques • use databases • analyse pipeline database information • create data models • debug software • interpret technical requirements • use software design patterns • use software libraries • utilise computer-aided software engineering (CASE) tools • adapt to changes in technological development plans • design user interfaces • use markup languages • provide technical documentation
Responsibilities and autonomy	The student has the ability to work independently in order to: • develop an analytical approach • take a proactive approach • develop strategies to solve problems • be open minded • coordinate engineering teams

8. Discipline objective (as results from the *key competences gained*)

8.1 General objective	Introducing students to data mining and data analysis processes; Knowledge of the techniques, algorithms and methods that can be used; Solutions identification and design for data analysis based on a given context
8.2 Specific objectives	Knowledge regarding data analysis methods and tools; Data processing; Determination and optimization of parameters; Methods for knowledge extraction

9. Contents

9.1 Lectures	Hours	Teaching methods	Notes
Introduction. What and why data analysis	1	Lectures using blackboard and projector; involving students in debate	
Advantages and challenges of data analytics	1		
Data analysis types and methods	1		
Data preparation. Dimensionality reduction	1		
Comparison, trend and ranking	1		
Variance, contribution and frequency	1		
Correlation and Pareto analysis	1		
Predictive data analysis	1		
Network analysis	1		
Data Visualisation	1		
Extracting knowledge through ML	1		
Time Series Analysis	1		
Framework for data analysis	1		

Realtime data	1		

Bibliography:

- Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. http://www.mmds.org/
- Hand, D.; Mannila, H.; Smyth, P. Principles of data mining. MIT Press, 2001. ISBN: 026208290X
- Spence, R, Pearson/Prentice Hall Information visualisation: design for interaction. 2007. ISBN: 9780132065504
- Domain articles

9.2 Applications - Seminars/Laboratory/Project	Hours	Teaching methods	Notes
Introduction to Data Analytics. Seminar/Laboratory topic selection and practical activity planning.	2	Oral presentation using slides, discutions (Q&A). Using multimedia tools, interactive teaching tools. Using specific software for data analysis	
Synthetic data generation	2		
Stream processing & Data storage	2		
Implementation of a data analysis pipeline – data acquisition, dataset description.	2		
Implementation of a data analysis pipeline – data cleaning and pre-processing.	2		
Hardware accelerators for data analytics	2		
Energy-efficient data-processing pipelines	2		
Implementation of a data analysis pipeline – data analysis procedures.	2		
Implementation of a data analysis pipeline – data/result visualization.	2		
Data analysis applications	2		
Evaluation of data analysis pipelines	2		
Implementation of a data analysis pipeline – Experiments	2		
Implementation of a data analysis pipeline – Final presentation	2		
Evaluation of activities.	2		

Bibliography

- Mohapatra, Shubhankar, et al. "Differentially Private Data Generation with Missing Data." *Proceedings of the VLDB Endowment* 17.8 (2024): 2022-2035.
- Gulisano, Vincenzo, and Alessandro Margara. "Aggregates are all you need (to bridge stream processing and Complex Event Recognition)." *Proceedings of the 18th ACM International Conference on Distributed and Event-based Systems*. 2024.
- Khelifati, Abdelouahab, et al. "TSM-bench: Benchmarking time series database systems for monitoring applications." *Proceedings of the VLDB Endowment* 16.11 (2023): 3363-3376.
- Kiefer, Martin, et al. "Optimistic data parallelism for fpga-accelerated sketching." *Proceedings of the VLDB Endowment* 16.5 (2023): 1113-1125.
- Herzog, Benedict, et al. "GreenPipe: Energy-Efficient Data-Processing Pipelines for Resource-Constrained Systems.", EWSN 2024
- Böther, Maximilian, et al. "Modyn: Data-Centric Machine Learning Pipeline Orchestration." *Proceedings of the ACM on Management of Data* 3.1 (2025): 1-30.
- Diao, Yiqun, et al. "OEBench: Investigating Open Environment Challenges in Real-World Relational Data Streams." *Proceedings of the VLDB Endowment* 17.6 (2024): 1283-1296.

**Se vor preciza, după caz: tematica seminariilor, lucrările de laborator, tematica și etapele proiectului.*

10. Bridging course contents with the expectations of the representatives of the community, professional associations and employers in the field

The class content was aligned with other similar classes from renowned universities and newest domain articles

11. Evaluation

Activity type	Assessment criteria	Assessment methods	Weight in the final grade
Course	Exam	Written exam - summative	50%
Seminar	Presentation	Oral examination, report -	25%

		continuous	
Laboratory	Exercises	Oral examination, demo, report - continuous	25%
Project	-	-	-
Minimum standard of performance: Final grade > 5			

Date of filling in: 01.09.2025	**Responsible**	**Title First name Last name**	**Signature**
	Course	Prof.dr.eng. Gheorghe SEBESTYEN	
	Applications	Prof.dr.eng. Anca HÂNGAN	

Date of approval in the department 17.09.2025	Head of department, Prof.dr.eng. Rodica Potolea
Date of approval in the Faculty Council 19.09.2025	Dean, Prof.dr.eng. Vlad Mureșan