

FIȘA DISCIPLINEI

(Introducere in Big Data)

Anul universitar 2025-2026

1. Date despre program

1.1. Instituția de învățământ superior	Universitatea Babeș-Bolyai Cluj-Napoca
1.2. Facultatea	Facultatea de Matematică și Informatică
1.3. Departamentul	Departamentul de informatică
1.4. Domeniul de studii	Informatică
1.5. Ciclu de studii	Licență
1.6. Programul de studii / Calificarea	Inteligență Artificială
1.7. Forma de învățământ	Cu frecvență

2. Date despre disciplină

2.1. Denumirea disciplinei	Introducere în Big Data			Codul disciplinei	MLE5203		
2.2. Titularul activităților de curs	Lect. dr. Ioana Ciuciu						
2.3. Titularul activităților de seminar	Lect dr. Ioana Ciuciu						
2.4. Anul de studiu	2	2.5. Semestrul	4	2.6. Tipul de evaluare	C	2.7. Regimul disciplinei	DS

3. Timpul total estimat (ore pe semestru al activităților didactice)

3.1. Număr de ore pe săptămână	4	din care: 3.2. curs	2	3.3. seminar/ laborator/proiect	2
3.4. Total ore din planul de învățământ	56	din care: 3.5. curs	28	3.6 seminar/laborator/proiect	28
Distribuția fondului de timp pentru studiul individual (SI) și activități de autoinstruire (AI)					ore
Studiul după manual, suport de curs, bibliografie și notițe (AI)					24
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					24
Pregătire seminare/ laboratoare/ proiecte, teme, referate, portofolii și eseuri					32
Tutoriat (consiliere profesională)					6
Examinări					8
Alte activități					-
3.7. Total ore studiu individual (SI) și activități de autoinstruire (AI)				94	
3.8. Total ore pe semestru				150	
3.9. Numărul de credite				6	

4. Precondiții (acolo unde este cazul)

4.1. de curriculum	
4.2. de competențe	<ul style="list-style-type: none">Cunoștințe de bază de analiza datelor; preferabilCunoștințe de bază de vizualizare a datelor; preferabilCompetențe de programare

5. Condiții (acolo unde este cazul)

5.1. de desfășurare a cursului	
5.2. de desfășurare a seminarului/ laboratorului	

6.1. Competențele specifice acumulate¹

¹ Se poate opta pentru competențe sau pentru rezultatele învățării, respectiv pentru ambele. În cazul în care se alege o singură variantă, se va șterge tabelul aferent celeilalte opțiuni, iar opțiunea păstrată va fi numerotată cu 6.

Competențe profesionale/esențiale	<ul style="list-style-type: none"> • gestioneaza proiecte de inginerie • creeaza modele de date
Competențe transversale	<ul style="list-style-type: none"> • lucrează în echipe • gândește analitic

6.2. Rezultatele învățării

Cunoștințe	<ul style="list-style-type: none"> • Absolventul cunoaște, înțelege și utilizează metode pentru reprezentarea, analiza și manipularea unor volume mari de date. • Absolventul este capabil sa descrie formal problemele abordate din diferite domenii, modelându-le ca probleme care se pot aborda cu tehnici din sfera inteligenței artificiale.
Aptitudini	<ul style="list-style-type: none"> • Absolventul este capabil să aplice algoritmi fundamentali de inteligență artificială pentru a rezolva probleme din lumea reală. • Absolventul este capabil să proiecteze și să implementeze sisteme software care folosesc metode ale inteligenței artificiale și să evalueze performanța acestora. • Absolventul are cunoștințele necesare pentru proiectarea, analiza și administrarea bazelor de date. • Absolventul este capabil să aplice șabloane arhitecturale, șabloane de proiectare și bunele practici în domeniu pentru a proiecta aplicații software de complexitate mare.
Responsabilități și autonomie	<ul style="list-style-type: none"> • Absolventul are abilitatea de a dezvolta, proiecta și crea noi aplicații, sisteme sau produse folosind bunele practici din domeniu. • Absolventul are cunoștințele necesare pentru selectarea și utilizarea procedurilor potrivite de instruire pentru a facilita procesul de asimilare a cunoștințelor.

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

7.1 Obiectivul general al disciplinei	<ul style="list-style-type: none"> • Înțelegerea și manipularea unor volume mari (și foarte mari) de date digitale disponibile în diverse formate (text, video, date financiare, medicale, structurate, nestructurate, semi-structurate)
7.2 Obiectivele specifice	<ul style="list-style-type: none"> • Permitea utilizării unor algoritmi noi, precum și a unor infrastructuri software și a unor metodologii în scopul manipulării (stocare, interogare, analiză) datelor de volum mare • Support de decizie asupra unor volume mari de date • Permitea creării de aplicații și servicii pentru diverse domenii de aplicare bazate pe rezultatele analizei de date de volum mare și foarte mare.

8. Conținuturi

8.1 Curs	Metode de predare	Observații
1. Introducere în Data Science și Big Data – partea I	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Principalele concepte ale Data Science, procesul Data Science, exemple, studii de caz
2. Introducere în Data Science și Big Data – partea II	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Tipuri de date, surse de date, disponibilitatea datelor, principalele concepte Big Data, tehnologii emergente, exemple de studii de caz
3. Standarde industriale pentru proiecte de Data Mining	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Medologia proiectelor din domeniul Data Science (CRISP-DM)
4. Arhitectura Big Data	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Baze de date tradiționale versus sisteme/arhitecturi Big Data
5. Procesarea de tip Batch (<i>Batch Layer</i>)	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Arhitectura Lambda - un model pentru sisteme Big Data, studii de caz și exemple
6. Procesarea datelor de tip Batch cu Apache Spark	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Stocarea datelor Big Data, modelul de date pentru Big Data, ecosistemul Hadoop Procesarea de tip batch, tehnologii. Exemple si exercitii folosind Apache Spark.
7. <i>Serving Layer</i>	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Specificatii, metrici de performanta, problema normalizarii/denormalizarii, tehnologii
8. Procesarea datelor in timp real (<i>Speed Layer</i>)	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Calculul și stocarea de view-uri in timp real (<i>real time views</i>), actualizări în timp real, tehnologii
9. Ingestia datelor	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Definiții și considerente de proiectare, ingestia de tip <i>batch</i> , ingestia datelor în timp real, tehnologii
10. Soluții NoSQL pentru Big Data	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Baze de date NoSQL, Modele de date NoSQL, tutorial
11. Provocari etice in Big Data	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația 	Provocari in dezvoltarea si folosirea aplicatiilor Big Data in ceea ce priveste (i) securitatea datelor; (ii) bias-ul algorithmic; (iii) transparenta; si (iv) implicarile etice si sociale
12. Vizualizarea datelor	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Principiile vizualizării științifice ale datelor, exemple, tehnologii
13. Studii de caz Big Data	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Prezentarea unor studii de caz din lumea reală
14. Prezentarea unui eseu de cercetare din domeniul Big Data	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația 	Prezentarea eseurilor individuale de către studenți

Bibliografie

Marz, N., & Warren, J. (2015). *Big Data. Principles and Best Practices of scalable real-time systems*. Manning Publications

Cielen, D., Meysman, A.D.B., & Ali, M. (2016). *Introducing Data Science. Big Data, machine learning, and more, using Python tools*. Manning Publications

Grus, J. (2019). *Data Science from Scratch: First Principles with Python*. O'Reilly Media, Inc.

Damji, J.S., Wenig, B., Das, T., & Lee, D. (2020). *Learning Spark*. O'Reilly Media, Inc.

Zečević, P., Bonaći, M. (2017). *Spark in Action*, Manning Publications

Perrin, J.G. (2020). *Spark in Action*, 2nd Ed., Manning Publications

Zelenin, A., Kropp, A. (2025). *Apache Kafka in Action*, Manning Publications

Sadalage, P., Fowler, M. (2013). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Pearson Education, Inc.

Banker, K., Bakkum, P., Verch, S., Garrett, D. and Hawkins, T. (2016). *MongoDB in Action*, Second Edition, Manning Publications

Borucki, A. (2024). *MongoDB in Action*, 3rd Ed., Manning Publications

Agneeswaran, V. (2014). *Big Data Analytics Beyond Hadoop*. Pearson Education

White, T. (2009). *Hadoop: The Definitive Guide*. O'Reilly

Holmes, A. (2015). *Hadoop in Practice*, 2nd ed., Manning Publications

McCallum, Q. E. (2012). *Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work*. O'Reilly

Grigorev, A. (2021). *Machine Learning Bookcamp*, Manning Publications

Rioux, J. (2022). *Data Analysis with Python and PySpark*, Manning Publications

Khalil, M. (2025). *Effective Data Analysis*, Manning Publications

8.2 Seminar / laborator	Metode de predare	Observații
<p>Proiect de semestru organizat cu echipe a cate 2-3 studenti (in functie de cerinte si de echipamentul disponibil, de ex., senzori inteligenti)</p> <p>Munca in echipa va fi autonoma (focusul pe creativitate si gandire critica)</p> <p>Vor fi furnizate tutoriale tehnice pentru a sustine munca studentilor in jurul celor mai importante aspecte ale stocarii si procesarii datelor (Tutorial Hadoop shell, Tutorial PySpark, Tutorial Spache Sqoop, Tutorial NoSQL, etc.)</p>	<p>Invatare bazata pe rezultate de cercetare</p> <p>Invatare sustinuta de tutoriale</p> <p>Munca in echipa</p> <p>Rezolvarea unor probleme legate de procesarea datelor Big Data</p>	<p>Echipele vor fi monitorizate de catre responsabilul de seminar/laborator prin intermediul unei platforme de proiect (in MS Teams).</p> <p>Seminarul se ține din două în două săptămâni și are două ore.</p>

Bibliografie (la fel ca si la curs)

1. <http://mahout.apache.org/>
2. http://www.tutorialspoint.com/mahout/mahout_introduction.htm
3. <http://spark.apache.org/documentation.html>
4. <http://shark.cs.berkeley.edu/>
5. <http://spark.apache.org/>
6. <http://nosql-database.org/>
7. <https://www.mongodb.com/nosql-explained>

9. Coroborarea conținuturilor disciplinei cu așteptările reprezentanților comunității epistemice, asociațiilor profesionale și angajatori reprezentativi din domeniul aferent programului

- Colaborări cu diverse inițiative locale și la nivel european: instituții naționale, industria IT, European Data Science Academy (EDSA, <https://edsa-project.eu/>), proiecte europene, European Federation of Data Driven Innovation Hubs (EUHubs4Data, <https://euhubs4data.eu/>), Incubator of Trusted B2B Data Sharing ecosystems of collaborating SMEs linked to Digital Innovation Hubs (i4Trust, <https://i4trust.org/>), REACH EuRopEAn incubator for trusted and secure data value Chains (<https://www.reach-incubator.eu/>), Big Data for Next Generation Energy (BD4NRG, <https://www.bd4nrg.eu/>), LETHE (<https://cordis.europa.eu/project/id/101017405>), FARE (<https://cordis.europa.eu/project/id/853566>), Human Brain Project (<https://www.humanbrainproject.eu/en/>), SoBigData (<http://project.sobigdata.eu/>), etc.
- Colaborări cu industria IT locală: sub formă de cursuri invitate, studii de caz reale (industriale), propuneri de subiecte pentru proiectul de semestru, echipament (e.g., senzori inteligenți).
- Colaborări cu alte programe de studiu (ex., cu Master in Bioinformatica de la Facultatea de Biologie) in cadrul proiectului de semestru sau cu studenti si profesori de la alte facultati si universitati (ex., proiecte colaborative, cursuri invitate)

10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere din nota finală
10.4 Curs	- familiarizarea cu principalele concept din domeniu - abilitatea de a aplica aceste principii în cazuri reale de utilizare	Examen scris/ Evaluarea eseului de cercetare	50%
10.5 Seminar/laborator	- abilitatea de a propune soluții creative viabile unor provocări big data din lumea reală - gândire critică - munca de cercetare individuală/în echipă	Proiect de semestru/Tutoriale	50%
10.6 Standard minim de performanță			
<ul style="list-style-type: none">• Minim nota 5 (pe o scală de la 1 la 10) este necesară pentru examenul scris, munca practică și eseul de cercetare.• Prezența la laborator este obligatorie în proporție de 90%, conform deciziei Consiliului Departamentului de Informatică (http://www.cs.ubbcluj.ro/wp-content/uploads/Hotarare-CDI-15.03.2017.pdf).			

11. Etichete ODD (Obiective de Dezvoltare Durabilă / Sustainable Development Goals)²

Nu se aplică.

Data completării:
15.04.2025

Semnătura titularului de curs
Lect.dr.Ioana Ciuciu

Semnătura titularului de seminar
Lect.Dr.Ioana Ciuciu



² Păstrați doar etichetele care, în conformitate cu [Procedura de aplicare a etichetelor ODD în procesul academic](#), se potrivesc disciplinei și ștergeți-le pe celelalte, inclusiv eticheta generală pentru *Dezvoltare durabilă* - dacă nu se aplică. Dacă nicio etichetă nu descrie disciplina, ștergeți-le pe toate și scrieți "Nu se aplică".

Data avizării în departament:

...

Semnătura directorului de departament

Conf.dr. Adrian STERCA