

FIȘA DISCIPLINEI

Procesare si Aplicatii 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 Biologie și Geologie
1.3. Departamentul	Departamentul de Biologie Moleculară și Biotehnologie
1.4. Domeniul de studii	Biologie
1.5. Ciclul de studii	Master
1.6. Programul de studii / Calificarea	Bioinformatică
1.7. Forma de învățământ	Cu frecvență

2. Date despre disciplină

2.1. Denumirea disciplinei		Procesare și Aplicații Big Data				Codul disciplinei	MME8158
2.2. Titularul activităților de curs			Lect. Dr. Ioana-Georgiana Ciuciu				
2.3. Titularul activităților de seminar			Lect. Dr. Ioana-Georgiana Ciuciu				
2.4. Anul de studiu	2	2.5. Semestrul	3	2.6. Tipul de evaluare	E	2.7. Regimul disciplinei	Obligativu

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

3.1. Număr de ore pe săptămână	3	Din2 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)					42
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					42
Pregătire seminare/ laboratoare/ proiecte, teme, referate, portofolii și eseuri					41
Tutoriat (consiliere profesională)					10
Examinări					4
Alte activități: comunicarea bi-direcționala cu responsabiul de curs/seminar					5
3.7. Total ore studiu individual (SI) și activități de autoinstruire (AI)				144	
3.8. Total ore pe semestru				200	
3.9. Numărul de credite				8	

4. Precondiții (acolo unde este cazul)

4.1. de curriculum	
4.2. de competențe	- Cunostinte de baza in data analytics - Cunostinte de baza in vizualizarea datelor - Abilitati de programare

5. Condiții (acolo unde este cazul)

5.1. de desfășurare a cursului	Proiector
5.2. de desfășurare a seminarului/ laboratorului	- Laborator cu calculatoare - Software big data instalat - Mediu de programare intr-un limbaj de nivel inalt

6.1. Competențele specifice acumulate¹

Competențe profesionale/esențiale	• Capacitatea de realizare de programe performante folosind programarea paralelă și distribuită • Modelare și rezolvarea eficientă de probleme din lumea reală
Competențe transversale	• Abilități de muncă în echipă, cu preluarea diferitelor roluri de execuție și conducere pentru realizarea unor proiecte • Abilități de comunicare profesională: descrierea clară, concisă, verbală și în scris, a rezultatelor profesionale, abilități de negociere

6.2. Rezultatele învățării

Cunoștințe	• Studentul știe să gestioneze cantități (extrem de) mari de date digitale în diverse formate (text, video, financiar, medical etc.) • Studentul cunoaște paradigmele de bază ale programării paralele • Studentul dobândește cunoștințele fundamentale care permit paralelizarea și rezolvarea unor probleme mari și complexe pe sisteme scalabile
Aptitudini	• Studentul este capabil să utilizeze algoritmi noi, infrastructuri software și metodologii în scopul prelucrării (stocării, achiziției/preluării, analizei) a unor cantități mari de date • Studentul este capabil să dezvolte aplicații și servicii pentru diverse domenii de afaceri pe baza analizei volumelor mari de date
Responsabilități și autonomie	• Studentul gestionează un flux de lucru și relaționează în echipă, ia decizii și gestionează situații neprevăzute, dezvoltă idei creative și tehnici inovative • Studentul dezvoltă și promovează strategii și practici de muncă eficiente, un stil și o conduită profesională exemplară, respectând valorile și principiile eticii și deontologiei profesionale • Studentul își dezvoltă capacitatea de a transla cunoștințele academice în context profesional, economic, social și etic. • Studentul utilizează strategii, metode și tehnici eficiente de învățare pe tot parcursul vieții, în vederea (auto)formării, (auto)dezvoltării personale și profesionale continue

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

7.1 Obiectivul general al disciplinei	• Gestiunea unor volume (extrem de) mari de date digitale în formate diferite (text, video, financiar, medical, etc.)
7.2 Obiectivele specifice	• Folosirea unor algoritmi, infrastructuri software și metodologii inovatoare pentru procesarea (memorare, extragere, analiză) unor volume mari de date • Oferirea de suport decizional peste volume largi de date • Crearea unor aplicații și servicii pentru diferite domenii de business pe baza rezultatelor analizei big data.

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.

8. Conținuturi

8.1 Curs	Metode de predare	Observații
1. Introducere in Data Science si Big Data	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Principalele concepte ale Data Science, procesul Data Science, exemple, studii de caz, tipuri de date, disponibilitatea datelor, principalele concepte Big Data, tehnologii emergente, exemple
2. Standarde Industriale pentru Proiecte de Data Mining	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Metodologia proiectelor din domeniul Data Science (CRISP-DM)
3. Arhitectura Big Data	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Baze de date tradiționale versus sisteme/arhitecturi Big Data, Arhitectura Lambda - un model pentru sisteme Big Data, studii de caz și exemple
4. Batch Layer	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Stocarea datelor Big Data, modelul de date pentru Big Data, ecosistemul Hadoop
5. Serving Layer	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Cerinte, metrice de performanta, problema normalizarii/denormalizarii, instrumente
6. Spark pentru procesarea datelor - partea I	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Procesarea datelor istorice folosind Apache Spark. Exemple
7. Speed Layer - part I	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Calculul și stocarea de view-uri in timp real (real time views), actualizări în timp real, tehnologii
8. Spark pentru procesarea datelor - partea II	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Procesarea datelor in timp real folosind Apache Spark. Exemple
9. Ingestia datelor	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Definiții și considerente de proiectare, ingestia de tip batch, ingestia datelor în timp real, tehnologii
10. Soluții NoSQL pentru Big Data	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Baze de date NoSQL, Modele de date NoSQL, tutorial
11. Provocări etice in Big Data	• Expunerea interactivă • Explicația • Exemple • Studii de caz	Provocări in dezvoltarea și folosirea aplicațiilor Big Data in ceea ce privește (i) securitatea datelor; (ii) bias-ul algorithmic;

	• Conversația	(iii) transparenta; si (iv) implicarile etice si sociale
12. Studii de caz Big Data	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Prezentarea unor studii de caz din lumea reală
13. Prezentarea unui eseu de cercetare din domeniul Big Data	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Prezentarea eseurilor individuale de către studenți
14. Prezentarea unui eseu de cercetare din domeniul Big Data	• Expunerea interactivă • Explicația • Exemple • Studii de caz • Conversația	Prezentarea eseurilor individuale de către studenți
Bibliografie Marz, N., & Warren, J. (2015). <i>Big Data. Principles and Best Practices of scalable real-time systems</i>. Manning Publications Cielen, D., Meysman, A.D.B., & Ali, M. (2016). <i>Introducing Data Science. Big Data, machine learning, and more, using Python tools</i>. Manning Publications Grus, J. (2019). <i>Data Science from Scratch: First Principles with Python</i>. O'Reilly Media, Inc. Damji, J.S., Wenig, B., Das, T., & Lee, D. (2020). <i>Learning Spark</i>. O'Reilly Media, Inc. Zečević, P., Bonaći, M. (2017). <i>Spark in Action</i>, Manning Publications Perrin, J.G. (2020). <i>Spark in Action</i>, 2nd Ed., Manning Publications Zelenin, A., Kropp, A. (2025). <i>Apache Kafka in Action</i>, Manning Publications Sadalage, P., Fowler, M. (2013). <i>NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence</i>. Pearson Education, Inc. Banker, K., Bakkum, P., Verch, S., Garrett, D. and Hawkins, T. (2016). <i>MongoDB in Action</i>, Second Edition, Manning Publications Borucki, A. (2024). <i>MongoDB in Action</i>, 3rd Ed., Manning Publications Agneeswaran, V. (2014). <i>Big Data Analytics Beyond Hadoop</i>. Pearson Education White, T. (2009). <i>Hadoop: The Definitive Guide</i>. O'Reilly Holmes, A. (2015). <i>Hadoop in Practice</i>, 2nd ed., Manning Publications McCallum, Q. E. (2012). <i>Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work</i>. O'Reilly Grigorev, A. (2021). <i>Machine Learning Bookcamp</i>, Manning Publications Rioux, J. (2022). <i>Data Analysis with Python and PySpark</i>, Manning Publications Khalil, M. (2025). <i>Effective Data Analysis</i>, Manning Publications		
8.2 Seminar / laborator	Metode de predare	Observații
Proiect de semestru organizat cu echipe a cate 2-3 studenti (in functie de cerinte si de echipamentul disponibil, de ex., senzori inteligenti)	Invatare bazata pe rezultate de cercetare	Echipele vor fi monitorizate de catre responsabilul de seminar/laborator prin intermediul unei platforme de proiect (in MS Teams).
	Invatare sustinuta de tutoriale	
	Munca in echipa	

Munca in echipa va fi autonoma (focusul pe creativitate si gandire critica)	Rezolvarea unor probleme legate de procesarea datelor Big Data	Seminarul se ține din două în două săptămâni și are două ore.
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.) 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). • Colaborari cu alte programe de studou (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	50%
10.6 Standard minim de performanță			
• 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 apr. 25

Semnătura titularului de curs
Lect.dr. Ioana-Georgiana CIUCIU

Semnătura titularului de seminar
Lect.dr. Ioana-Georgiana CIUCIU

Data avizării în departament:
...

Semnătura directorului de departament
Conf.dr. Adrian STERCA

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."