

FIȘA DISCIPLINEI

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 studiu / Calificarea	Bioinformatică

2. Date despre disciplină

2.1 Denumirea disciplinei (ro)/(en)	Procesare și Aplicații Big Data Big Data Processing and Applications		
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	Obligativiu
2.8 Codul disciplinei	MME8158		

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	2
3.4 Total ore din planul de învățământ	56	Din care: 3.5 curs	28	3.6 seminar/laborator	28
Distribuția fondului de timp:					ore
Studiul după manual, suport de curs, bibliografie și notițe					42
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					42
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					46
Tutoriat					10
Examinări					4
Alte activități:					-
3.7 Total ore studiu individual	140				
3.8 Total ore pe semestru	196				
3.9 Numărul de credite	8				

4. Precondiții (acolo unde este cazul)

4.1 de curriculum	
4.2 de competențe	<ul style="list-style-type: none"> • Cunoștințe de baza în data analytics • Cunoștințe de baza în vizualizarea datelor • Abilități de programare

5. Condiții (acolo unde este cazul)

5.1 De desfășurare a cursului	<ul style="list-style-type: none"> • Proiector
5.2 De desfășurare a seminarului/laboratorului	<ul style="list-style-type: none"> • Laborator cu calculatoare • Software big data instalat • Mediu de programare într-un limbaj de nivel înalt

6. Competențele specifice acumulate

Competențe profesionale	<p>C3.2 Identificarea și explicarea modelelor informatice de baza adecvate domeniului de aplicare</p> <p>C3.3 Utilizarea modelelor și instrumentelor informatice și matematice pentru rezolvarea problemelor specifice domeniului de aplicare</p> <p>C3.4 Analiza datelor și a modelelor</p> <p>C5.3 Utilizarea metodologiilor și mediilor de proiectare a bazelor de date pentru probleme particulare</p>
Competențe transversale	<p>CT1 Aplicarea regulilor de muncă organizată și eficientă, a unor atitudini responsabile față de domeniul didactic-științific, pentru valorificarea creativă a propriului potențial, cu respectarea principiilor și a normelor de etică profesională</p> <p>CT2 Desfășurarea eficientă a activităților organizate într-un grup inter-disciplinar și dezvoltarea capacităților empatică de comunicare inter-personală, de relaționare și colaborare cu grupuri diverse</p> <p>CT3 Utilizarea unor metode și tehnici eficiente de învățare, informare, cercetare și dezvoltare a capacităților de valorificare a cunoștințelor, de adaptare la cerințele unei societăți dinamice și de comunicare în limba română și într-o limbă de circulație internațională</p>

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

7.1 Obiectivul general al disciplinei	<ul style="list-style-type: none"> • Gestiunea unor volume (extrem de) mari de date digitale în formate diferite (text, video, financiar, medical, etc.)
7.2 Obiectivele specifice	<ul style="list-style-type: none"> • Folosirea unor algoritmi, infrastructuri software și metodologii inovatoare pentru procesarea (memorare, extragere, analiză) unor volume mari de date • Oferirea de suport în decizii peste volume largi de date • Crearea unor aplicații și servicii pentru diferite domenii de business pe baza rezultatelor analizei big data.

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. Standarde industriale pentru proiecte de Data Mining	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Metodologia proiectelor din domeniul Data Science (CRISP-DM)
3. Arhitectura Big Data	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Baze de date tradiționale versus sisteme/arhitecturi Big Data, Arhitectura Lambda - un model pentru sisteme Big Data, studii de caz și exemple

4. Procesarea datelor de tip Batch (<i>Batch Layer</i>)	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Stocarea datelor Big Data, modelul de date pentru Big Data, ecosistemul Hadoop
5. Serving Layer – partea I	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Cerinte, metrice de performanta, problema normalizarii/denormalizarii, instrumente
6. Serving Layer – partea II		
7. Procesarea datelor in timp real (<i>Speed Layer</i>) – partea 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
8. Procesarea datelor în timp real (<i>Speed Layer</i>) – partea II		
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. Vizualizarea datelor	<ul style="list-style-type: none"> • Expunerea interactivă • Explicația • Conversația • Demonstrația didactică 	Principiile vizualizării științifice ale datelor, exemple, tehnologii
12. 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
13. 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
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.

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

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

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

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

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

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

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.</p> <p>Seminarul se ține din două în două săptămâni și are două ore.</p>

Bibliografie

Identica cu bibliografia pentru curs

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, ca de exemplu 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>), the 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).

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	Examen scris/ Evaluarea eseului de cercetare	50%

	- abilitatea de a aplica aceste principii în cazuri reale de utilizare		
10.5 Laborator	<p>- abilitatea de a propune soluții creative viabile unor provocări big data din lumea reală</p> <p>-abilitatea de a consuma (interoga, analiza) date de tip Big Data cu scopul de a extrage informații relevante pentru diverse studii de caz și domenii de aplicare</p> <p>- demonstrarea gândirii critice</p> <p>- abilitatea de a realiza sarcini individual și în echipă</p>	Proiect de semestru	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). 			

Data completării

19.06.2024

Semnătura titularului de curs

Lect. Dr. Ioana-Georgiana Ciuciu

Semnătura titularului de seminar

Lect. Dr. Ioana-Georgiana Ciuciu

Data avizării în departament

.....19.06.2024.....

Semnătura directorului de departament

Conf. Dr. Sterca Adrian