



FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea “Alexandru Ioan Cuza” din Iași
1.2 Facultatea	Facultatea de Informatică
1.3 Departamentul	Informatică
1.4 Domeniul de studii	Informatică
1.5 Ciclul de studii	Studii universitare de masterat
1.6 Programul de studii / Calificarea	Ingineria Sistemelor Software(OB), Optimizare Computațională (OP), Lingvistică Computațională (OP)

2. Date despre disciplină

2.1 Denumirea disciplinei	Analiza bazelor mari de date						
2.2 Titularul activităților de curs	Conf. dr. Mihaela BREABĂN						
2.3 Titularul activităților de seminar	Conf. dr. Mihaela BREABĂN						
2.4 An de studiu	1,2	2.5 Semestru	2	2.6 Tip de evaluare	E	2.7 Regimul disciplinei*	OB

* OB – Obligatoriu / OP – Opțional

3. Timpul total estimat (ore pe semestru și activități 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
Studiu după manual, suport de curs, bibliografie și altele					28
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					20
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					56
Tutoriat					
Examinări					4
Alte activități					
3.7 Total ore studiu individual					104
3.8 Total ore pe semestru					164
3.9 Număr de credite					8

4. Precondiții (dacă este cazul)

4.1 De curriculum	Baze de date, Structuri de date, Proiectarea algoritmilor, Programare orientată obiect, Învățare automată, Ingineria programării
4.2 De competențe	Crearea, manipularea și interogarea bazelor de date relaționale; Programare Java

5. Condiții (dacă este cazul)

5.1 De desfășurare a cursului	--
5.2 De desfășurare a seminarului/laboratorului	--



6. Competențe specifice acumulate

Competențe profesionale	La finalizarea cu succes a acestei discipline, studenții C1. Vor cunoaște principalele modele de baze de date, vor fi capabili să ia decizii asupra celui mai potrivit sistem de gestiune de baze de date (relațional sau din grupul NoSQL) într-o situație dată sau să asigure interoperabilitatea mai multor sisteme C2. Vor cunoaște tehnologii, paradigme și algoritmi eficienți dedicați analizei unor volume mari de date cu importanță aplicabilitate practică C3. Vor fi capabili să implementeze în mod eficient algoritmi de procesare distribuită a datelor
Competențe transversale	CT1. Integrarea unui modul de gestionare eficientă a datelor în diverse aplicații software CT2. Eficientizarea procesării unui volum mare de date cu și a analizei lor cu paradigme ale învățării automate utilizând baze de date NoSQL și MapReduce CT3. Procesarea datelor nestructurate în cadrul aplicațiilor WEB

7. Obiectivele disciplinei (din grila competențelor specifice acumulate)

7.1 Obiectivul general	Deprinderea abilităților de gestionare și procesare eficientă a volumelor mari de date în medii distribuite, asigurând interoperabilitatea sistemelor de gestiune de baze de date relationale și NoSQL și implementând algoritmi eficienți de procesare distribuită
7.2 Obiectivele specifice	O1. Compararea și exemplificarea principalelor modele de baze de date O2. Înțelegerea particularităților sistemelor de fișiere distribuite: teorema CAP, proprietățile BASE, consistența asimptotică O3. Însușirea modelului de procesare distribuită a datelor conform paradigmei MapReduce O4. Însușirea unor algoritmi de analiză a datelor și predicție O5. Integrarea a diverse sisteme de gestiune și procesare de date în aplicații reale

8. Conținut

8.1	Curs	Metode de predare	Observații (ore și referințe bibliografice)
1.	Modele și sisteme de gestiune de baze de date: istoric, particularități, proprietăți ACID vs. BASE, teorema CAP în sisteme distribuite	Expunere, problematizare, discuții	2
2.	Modele de calcul paralel; eficiență; paradigma Map Reduce	Expunere	2
3.	Ecosistemul Hadoop. Componentele de bază: sistemul de fișiere distribuit (HDFS), frameworkul Map-Reduce	Expunere	2
4.	Hadoop Map-Reduce: fazele dezvoltării de aplicații în mediu distribuit	Expunere	2



5.	Map-Reduce: implementări eficiente pentru algoritmi de bază (operatori relaționali, înmulțiri matriceale, indexare)	Expunere, problematizare	2
6.	Hadoop – integrare cu alte sisteme într-un centru de date.	Expunere, Exemplificare	2
7.	Suport pentru modelul relational (Pig, Hive, Impala).	Expunere	2
8.	Map-Reduce: eficiență	Problematizare, discuții	2
9.	Apache Spark	Expunere	2
10.	Motoare de căutare: algoritmul Page Rank	Expunere	2
11.	Analiza structurii rețelelor sociale: clustering în grafuri	Expunere	2
12.	Sisteme de recomandare (1)	Expunere	2
13.	Sisteme de recomandare (2)	Expunere	2
14.	Tehnici eficiente pentru căutare bazată pe similaritate	Expunere	2

Bibliografie

- Avi Silberschatz Henry F. Korth S. Sudarshan: *Database System Concepts*. McGraw-Hill Science/Engineering/Math; 4th edition
- Anand Rajaraman, Jeffrey David Ullman: *Mining of Massive Datasets*, Cambridge University Press (November 30, 2011)
- Tom White: *Hadoop: The definitive guide*, O'Reilly Media, Inc., 2012.
- Eric Samme: *Hadoop operations*, O'Reilly Media, Inc., 2012.

8.2	Seminar / Laborator	Metode de predare	Observații (ore și referințe bibliografice)
1.	Lucrul cu sistemul de fișiere distribuit Hadoop, lansarea unui job Map-Reduce în manieră pseudo-distribuită	Exercitii, Metode interactive	2
2.	Scrierea unui program Map-Reduce; unit-testing	Experiment, Lucru individual, Metode interactive	2
3.	Clasa ToolRunner; Combiners; Counters; mecanisme de logging	Experiment, Lucru individual, Metode interactive	2
4.	Implementarea unui <i>Partitioner</i> , definirea de noi clase <i>WritableComparable</i> , utilizarea fișierelor <i>SequenceFile</i>	Experiment, Lucru individual, Metode interactive	2
5.	Implementări Map-Reduce: Inverted Index, Words Co-Occurrence. Importarea datelor (Sqoop, Flume). Definirea de fluxuri de lucru (Oozie)	Experiment, Lucru individual, Metode interactive	2



6.	Suport pentru modelul relational: Apache Pig	Experiment, Lucru individual, Metode interactive	2
7.	Suport pentru modelul relational: Apache Hive	Experiment, Lucru individual, Metode interactive	2
8.	Evaluare intermediară		2
9.	Apache Spark - MLlib	Experiment, Lucru individual, Metode interactive	2
10.	Proiect: definirea cerințelor, planificare	Problematizare, Lucru în echipă	2
11.	Proiect: implementare	Experiment, Lucru în echipă	2
12.	Proiect: evaluare a progresului	Prezentări, discuții	2
13.	Proiect: implementare	Experiment, Lucru în echipă	2
14.	Proiect: evaluare finală	Evaluare finală	2

Bibliografie

Tutoriale online – trimiteri de pe pagina laboratorului (www.info.uaic.ro/~pmihaela/BDA)

9. Coroborarea conținutului disciplinei cu așteptările reprezentanților comunității, asociațiilor profesionale și angajatorilor reprezentativi din domeniul aferent programului

Conținutul disciplinei este pe deplin coroborat cu necesitățile angajatorilor din industria software și nu numai.

Cunoștințele dobândite în cadrul cursului sunt necesare procesării eficiente a unor volume mari de date cu algoritmi de data mining, abilitate necesară în cercetarea din multe domenii.

10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere în nota finală (%)
10.1 Curs	-înțelegerea caracteristicilor particulare modelelor de baze de date studiate -capacitatea de a proiecta algoritmi eficienți conform paradigmei Map Reduce -cunoașterea particularităților HDFS și a	Test scris în săptămâna a opta a semestrului	1/3 (33,3%)



	proiectelor studiate din cadrul <i>ecosistemul Hadoop</i>		
10.2 Seminar/ Laborator	- cunoașterea elementelor de bază din API-ul Hadoop Map-Reduce și a proiectelor înrudite (Pig, Hive, Oozie, Sqoop, Flume), rezolvarea de probleme de analiza date cu Apache Spark	Evaluare teme săptămânale pe parcursul primei jumătăți a semestrului	1/3 (33,3%)
10.3 Mixt (Curs+laborator)	Implementarea cu succes a unor algoritmi predați în a doua jumătate a semestrului la curs, în cadrul unui proiect amplu în echipă	Evaluarea progresului proiectului pe parcurs și a gradului de realizare a acestuia în săptămâna a 14-a a semestrului	1/3 (33,3%)
10.6 Standard minim de performanță			
Pentru promovare trebuie îndeplinite simultan următoarele criterii: <ul style="list-style-type: none">• Minim 50% din punctajul maxim la test• Minim 50% din punctajul total maxim la temele pe parcurs• Minim 50% din punctajul total maxim care însumează cele trei probe de evaluare Notele se stabilesc conform cu criteriile ECTS			

Data completării
februarie 2018

Titular de curs
Conf. dr. Mihaela Breabăn

Titular de seminar
Conf. dr. Mihaela Breabăn

Data avizării în departament

Director de departament